# HIGHER Mathematics

V.S. Shipachev

Mir Publishers
Moscow

# HIGHER Mathematics

V.S.Shipachev

**Higher Mathematics**

**В. С. Шипачев**

# ВЫСШАЯ МАТЕМАТИКА

Под редакцией
академика А. Н. Тихонова

«ВЫСШАЯ ШКОЛА» Москва

V. S. Shipachev

# HIGHER Mathematics

Edited by Academician
A. N. Tikhonov

Mir Publishers Moscow

Translated from the Russian
by Irene Aleksanova

First published 1988
Revised from the 1985 Russian edition

TO THE READER

Mir Publishers would be grateful for your comments on the content, translation and design of this book. We would also be pleased to receive any other suggestions you may wish to make.

Our address is:
Mir Publishers
2 Pervy Rizhsky Pereulok
I-110, GSP, Moscow, 129820
USSR

*На английском языке*

*Printed in the Union of Soviet Socialist Republics*

ISBN 5-03-000233-2

© Издательство «Высшая школа», 1985
© English translation, Mir Publishers, 1988

## FROM THE EDITOR

In this age of scientific and technological progress, the mathematical methods of research permeate every field of human activity. Allied sciences use a great deal of mathematics and raise new problems in the study of mathematics itself. Mathematics as a school discipline holds a firm place in the curricula of nonmathematical faculties of higher educational establishments. The study of mathematics develops logical thinking and provides necessary knowledge for understanding the most complicated problems facing the economy.

This textbook is based on *The Course in Higher Mathematics* by the same author, which was published by Moscow University Press and won recognition from many mathematicians; it was awarded a bronze medal at the Exhibition of Economic Achievement in Moscow in 1983.

For many years the author has lectured on higher mathematics at the Faculty of Geology of Moscow University. At the same time he worked at the courses, sponsored by Moscow University, for raising the scientific qualification of teachers of mathematics at secondary schools. He was awarded the titles of Excellent Worker of Public Education of the RSFSR and Excellent Worker of Education of the USSR. Shipachev's experience of teaching has left its mark on both the material selected and its exposition. Although the book covers a large range of material, the exposition is clear and compact. A large number of examples and problems are given to illustrate the basic concepts and theorems and will be conducive to the mastering by the students of the fundamentals of higher mathematics and show them how it can be used in solving applied problems.

The happy combination of simplicity and strictness of the exposition of the material and thoroughly selected examples will make it possible for students of nonmathematical faculties of higher educational establishments to use the book as their fundamental textbook.

In 1987 this book was awarded a diploma of honour of the Exhibition of Economic Achievement.

*Academician A.N. Tikhonov*

## PREFACE

The book gives the fundamentals of higher mathematics and can therefore be used both at universities and at higher technical schools where different divisions of higher mathematics are usually combined into a single course.

I have tried to present the material as completely, strictly and simply as possible. My aim was not only to convey a certain quantity of knowledge in higher mathematics but also to arouse an interest in mathematics in my readers, to broaden their views and to foster a mathematical culture.

The exposition is based on the principle of didactics, i.e. from the simple to the more complicated. Thus, for instance, the concept of a limit is first introduced for number sequences, then for functions of one variable, then for integral sums and, finally, for functions of several variables.

Real numbers are introduced axiomatically, which makes it possible to give the necessary knowledge of numbers in compact form and to proceed with the basic material in simpler manner.

The bulk of the material is devoted to mathematical analysis, which includes differential and integral calculus and the most important concept of higher mathematics, that of a function. This concept is studied as early as the course in elementary mathematics, but higher mathematics investigates it more completely and systematically. It should be pointed out that the function is defined in terms of the set and this corresponds to present fashion in teaching mathematics.

The textbook contains, in the main, theoretical material, and therefore for practical purposes various other problem books should be used.

It is my honour to express my deep gratitude to Academician A.N. Tikhonov and to Professors A.G. Sveshnikov, D.P. Kostomarov, N.M. Matveev and A.A. Shestakov for their active participation in the work on the book.

May I also convey my appreciation to Professor A.A. Babaev, Corresponding Member of the Academy of Sciences of the Azerbaijan-

ian SSR, Professor G.I. Chandirov and all the lecturers at the Chair of Mathematics in the Physics Faculty of the Azerbaijanian State University all of whom took part in the discussion of the manuscript and made valuable remarks.

My thanks go also to Professor V.F. Butuzov and Assistant Professor V.M. Govorov, who are talented mathematicians and lecturers and who made valuable contributions to the creation of this textbook.

*The author*

**NOTATION**

| | |
|---|---|
| $x \in X$ | The element $x$ belongs to the set $X$ |
| $x \notin X$ | The element $x$ does not belong to the set $X$ |
| $X \subset Y$ | The set $X$ is a subset of the set $Y$ |
| $X \not\subset Y$ | The set $X$ is not a subset of the set $Y$ |
| $\phi$ | An empty set |
| $\{x \mid P(x)\}$ | The set of numbers $x$ which possess the property $P(x)$ |
| $x = \max\{x_1, \ldots, x_n\}$ | $x$ is the maximum of the numbers $x_1, \ldots, x_n$ |
| $x = \min\{x_1, \ldots, x_n\}$ | $x$ is the minimum of the numbers $x_1, \ldots, x_n$ |
| $\forall x$ | For any $x$ |
| $\exists x$ | There is $x$ such that... |
| $(x, y)$ | The coordinates of a point on the plane, an ordered pair of numbers |
| $(x, y, z)$ | The coordinates of a point in space, an ordered triple of numbers |
| $\overrightarrow{AB}$ | A directed segment |
| $\overrightarrow{AB}$, $\mathbf{a}$ | A vector |
| $\lvert \overrightarrow{AB} \rvert$, $\lvert \mathbf{a} \rvert$ | The length of a vector |
| $\overrightarrow{AB} = \{X, Y, Z\}$ | The vector $\overrightarrow{AB}$ has coordinates $X, Y, Z$ |
| $\mathrm{proj}_u \overrightarrow{AB}$ | The projection of the vector $\overrightarrow{AB}$ on the $u$-axis |
| $\mathbf{a} \cdot \mathbf{b}$ | The scalar product of the vectors $\mathbf{a}$ and $\mathbf{b}$ |
| $\mathbf{a} \times \mathbf{b}$ | The vector product of the vectors $\mathbf{a}$ and $\mathbf{b}$ |

| | |
|---|---|
| $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})$ | The triple, or parallelepipedal, product of the vectors **a**, **b**, **c** |
| $[a, b]$ | A closed interval |
| $(a, b)$, $(a, +\infty)$, $(-\infty, b)$, $(-\infty, +\infty)$ | Open intervals |
| $[a, b)$, $(a, b]$, $[a, +\infty)$, $(-\infty, b]$ | Half-open or half-closed intervals |
| $(-\infty, +\infty)$ | The number line, the set of all real numbers |
| $(a-\delta, a+\delta)$ | The $\delta$-neighbourhood of the point $a$ |
| $\sup X$ | The least upper bound of the set $X$ |
| $\inf X$ | The greatest lower bound of the set $X$ |
| $\lvert x \rvert$ | The absolute value (modulus) of the number $x$ |
| $\operatorname{sgn} x$ | Signum function |
| $[x]$ | The integral part of the number $x$ |
| $\{x_n\}$ | A number sequence |
| $\sum_{n=1}^{n} a_n$ | A number series |
| $y=f(x)$, $f\colon x\to y$ | A function, a mapping |
| $\lim_{x\to x_0} f(x)$ | The limit of the function $f(x)$ as $x\to x_0$ |
| $\lim_{x\to x_0+} f(x)$ | The limit of the function $f(x)$ as $x\to x_0$ from the right, the right-hand limit |
| $\lim_{x\to x_0-} f(x)$ | The limit of the function $f(x)$ as $x\to x_0$ from the left, the left-hand limit |
| $\alpha(x)\sim\beta(x)$ | The infinitesimals $\alpha(x)$ and $\beta(x)$ are equivalent |
| $\alpha(x)=o(\beta(x))$ | The infinitesimal $\alpha(x)$ has the higher order of smallness as compared to the infinitesimal $\beta(x)$ |
| $\sup_X f(x)$ | The least upper bound of the function $f(x)$ on the set $X$ |
| $\inf_X f(x)$ | The greatest lower bound of the function $f(x)$ on the set $X$ |
| $f'(x)$, $y'(x)$ | The derivative of the function $y=f(x)$ at the point $x$ |
| $f'_+(x)$ | The right-hand derivative of the function $y=f(x)$ at the point $x$ |

| | |
|---|---|
| $f'_-(x)$ | The left-hand derivative of the function $y = f(x)$ at the point $x$ |
| $dy$ | The differential of the function $y = f(x)$ |
| $y'', y''', \ldots, y^{(n)}$ | The derivatives of the second, third, ..., $n$th order |
| $d^2y, d^3y, \ldots, d^ny$ | Differentials of the second, third, ..., $n$th order |
| $\Delta y, \Delta^2 y, \ldots, \Delta^n y$ | Differences of the first, second, ..., $n$th order |
| $z = f(x, y), z = f(M)$ | A function of two variables |
| $\lim\limits_{\substack{x\to x_0\\ y\to y_0}} f(x, y), \lim\limits_{M\to M_0} f(M)$ | The limit of the function $z = f(x, y) = f(M)$ at the point $M_0(x_0, y_0)$ |
| $z'_x, f'_x, \frac{\partial f}{\partial x}, \frac{\partial z}{\partial x}$ | A partial derivative of the function $z = f(x, y)$ with respect to $x$ |
| $\frac{\partial^2 z}{\partial x\, \partial y}, f''_{xy}(x, y)$ | Second-order mixed partial derivative of the function $z = f(x, y)$ |
| $\int f(x)\, dx$ | An indefinite integral of the function $f(x)$ |
| $\int\limits_a^b f(x)\, dx$ | A definite integral of the function $f(x)$ over the interval $[a, b]$ |
| $\iint\limits_G f(x, y)\, ds,$ $\iint\limits_G f(x, y)\, dx\, dy$ | A double integral of the function $f(x, y)$ over the domain $G$ |
| $\int\limits_{AB} f(M)\, dl,$ $\int\limits_{AB} f(x, y)\, dl$ | A line integral of the first kind of the function $f(x, y) = f(M)$ over the curve $AB$ |
| $\int\limits_{AB} P(x, y)\, dx$ | A line integral of the second kind of the function $P(x, y)$ over the curve $AB$ |
| $\oint\limits_L P(x, y)\, dx + Q(x, y)\, dy$ | A line integral around a closed contour $L$ |

| | |
|---|---|
| $\iiint\limits_V f(x, y, z)\, dv$, $\iiint\limits_V f(x, y, z)\, dx\, dy\, dz$ | A triple integral of the function $f(x, y, z)$ over the domain $V$ |
| $\iint\limits_S f(M)\, dS$, $\iint\limits_S f(x, y, z)\, dS$ | A surface integral of the first kind of the function $f(x, y, z) = f(M)$ over the surface $S$ |
| $\iint\limits_S R(M)\, dx\, dy$, $\iint\limits_S R(x, y, z)\, dx\, dy$ | A surface integral of the second kind of the function $R(x, y, z) = R(M)$ over the surface $S$ |
| grad z | The gradient of the function $z = f(x, y)$ |
| div **a** | The divergence of the vector field **a** $(M)$ |
| rot **a** | The rotation of the vector field **a** $(M)$ |
| $\nabla$ | The Hamiltonian operator |
| $\Delta$ | The Laplacian operator |

CONTENTS

From the Editor 5
Preface 6
Notation 8
Introduction 16

Part One. Analysis of Functions of One Variable 21

Chapter 1. Real Numbers 21

1.1.1. Sets. Notation. Logical Symbols 21
1.1.2. Real Numbers and Their Principal Properties 22
1.1.3. Geometric Representation of Real Numbers 25
1.1.4. Bounds of Number Sets 28
1.1.5. The Absolute Value of a Number 30

Chapter 2. The Limit of a Sequence 31

1.2.1. Number Sequences 32
1.2.2. Convergent Sequences 36
1.2.3. Monotone Sequences 42
1.2.4. Theorem on Nested Intervals 44

Chapter 3. Plane Analytic Geometry 45

1.3.1. Rectangular System of Coordinates 46
1.3.2. Simple Problems of Plane Analytic Geometry 47
1.3.3. Polar Coordinates 49
1.3.4. Transformation of Rectangular Coordinates 50
1.3.5. The Equation of a Plane Curve 53
1.3.6. First-Order Curves 55
1.3.7. Second-Order Curves 64
1.3.8. The General Equation of a Second-Order Curve 77

Chapter 4. Functions of One Variable 82

1.4.1. The Function 83
1.4.2. The Limit of a Function 87
1.4.3. Theorems on the Limits of Functions 92
1.4.4. Two Remarkable Limits 93
1.4.5. Infinitely Small and Infinitely Large Functions 96
1.4.6. Comparing Infinitely Small and Infinitely Large Functions 99
1.4.7. Continuity of a Function 102
1.4.8. Continuity of Some Elementary Functions 104
1.4.9. Classification of Points of Discontinuity of Functions 106
1.4.10. Principal Properties of Continuous Functions 107

1.4.11. Composite Functions 116
1.4.12. Inverse Functions 117

Chapter 5. Differentiation 121

1.5.1. The Derivative 121
1.5.2. The Differentiability of a Function 124
1.5.3. The Differential 126
1.5.4. The Rules of Differentiating the Sum, the Difference, the Product and the Quotient 128
1.5.5. Calculating the Derivatives of a Constant Function and a Power Function, of Trigonometric Functions and Logarithmic Functions 129
1.5.6. Theorem on the Derivative of an Inverse Function 132
1.5.7. Calculating the Derivatives of an Exponential Function and of Inverse Trigonometric Functions 133
1.5.8. The Rule of Differentiating a Composite Function 135
1.5.9. The Logarithmic Derivative. The Derivative of a Power Function with Any Real Exponent. Table of Derivatives of the Basic Elementary Functions 137
1.5.10. Higher-Order Derivatives and Differentials 139
1.5.11. The Parametric Representation of a Function and Its Differentiation 144

Chapter 6. Investigating Functions by Differentiation 146

1.6.1. Fundamental Theorems of Differential Calculus 147
1.6.2. Evaluating Indeterminate Forms. L'Hôpital's Rule 151
1.6.3. Taylor's Formula 155
1.6.4. Investigating the Behaviour of Functions and Constructing Graphs 161
1.6.5. Interpolation of Functions 173
1.6.6. Methods of Approximation of the Roots of Equations 178

Chapter 7. The Indefinite Integral 181

1.7.1. The Antiderivative and the Indefinite Integral 181
1.7.2. Principal Properties of the Indefinite Integral 183
1.7.3. Table of Fundamental Integrals 185
1.7.4. Principal Methods of Integration 185
1.7.5. Integration of Rational Functions 191
1.7.6. Integration of Irrational and Transcendental Functions 196

Chapter 8. The Definite Integral 202

1.8.1. Definition of the Definite Integral 202
1.8.2. The Conditions for the Existence of the Definite Integral 204
1.8.3. Integrability of Continuous and Certain Discontinuous Functions 210
1.8.4. Principal Properties of the Definite Integral 212
1.8.5. Evaluation of Integrals. Mean Value Formula 214
1.8.6. An Integral with a Variable Upper Limit 217
1.8.7. The Newton-Leibniz Formula 219
1.8.8. A Change of Variable in the Definite Integral 221
1.8.9. The Formula for Integration by Parts in the Definite Integral 222
1.8.10. Some Applications of the Definite Integral in Physics and Geometry 224
1.8.11. Improper Integrals 237
1.8.12. Approximation of Definite Integrals 243

**Part Two. Analysis of Functions of Several Variables** 251

**Chapter 9. Solid Analytic Geometry** 251

2.9.1. Rectangular Coordinate System in Space 251
2.9.2. The Vector 252
2.9.3. Linear Operations Involving Vectors and Their Principal Properties 256
2.9.4. Theorems on the Projections of Vectors 259
2.9.5. Representation of a Vector in Terms of Base Vectors 261
2.9.6. The Scalar Product of Vectors 262
2.9.7. The Vector Product of Vectors 265
2.9.8. The Triple Product of Three Vectors 270
2.9.9. Equations of a Surface and of a Curve 272
2.9.10. The Equation of a Cylindrical Surface 273
2.9.11. Equations of a Plane 275
2.9.12. The Equation of a Straight Line 279
2.9.13. Mutual Positions of a Straight Line and a Plane 283
2.9.14. Second-Order Surfaces 284

**Chapter 10. Elements of Higher Algebra** 291

2.10.1. Matrices 291
2.10.2. Determinants 295
2.10.3. Investigating the Behaviour of a System of Three First-Degree Equations with Three Unknowns 300
2.10.4. The Matrix Notation of a System of Linear Equations. The Inverse of a Matrix 305

**Chapter 11. The Concept, the Limit and the Continuity of a Function of Several Variables** 308

2.11.1. The Concept of a Function of Several Variables 308
2.11.2. Geometric Representation of a Function of Two Variables 310
2.11.3. The Limit of a Function of Two Variables 312
2.11.4. The Continuity of a Function of Two Variables 314

**Chapter 12. Partial Derivatives and the Differentiability of Functions of Several Variables** 318

2.12.1. Partial Derivatives 318
2.12.2. The Differentiability of a Function 319
2.12.3. Derivatives of Composite Functions 323
2.12.4. The Differential of a Function 326
2.12.5. The Directional Derivative. The Gradient 328
2.12.6. Higher-Order Partial Derivatives and Differentials 331
2.12.7. Taylor's Formula for a Function of Two Variables 335
2.12.8. The Extrema of a Function of Two Variables 337
2.12.9. The Method of Least Squares 341

**Chapter 13. Integration** 343

2.13.1. Double Integrals 343
2.13.2. Reducing a Double Integral to an Iterated Integral 347
2.13.3. A Change of Variables in a Double Integral 352
2.13.4. Some Geometric and Physical Applications of Double Integrals 355
2.13.5. Line Integrals 363
2.13.6. Green's Theorem 374

2.13.7. The Conditions for Independence of a Line Integral of the Path of Integration 376
2.13.8. Integration of Total Differentials 381
2.13.9. Some Applications of the Second-Kind Line Integrals 385
2.13.10. Triple Integrals 388
2.13.11. Surface Integrals 395
2.13.12. Ostrogradsky's Theorem 405
2.13.13. Stokes' Theorem 408
2.13.14. Scalar and Vector Fields 412

Part Three. Series. Differential Equations 423

Chapter 14. Series 423

3.14.1. Number Series 423
3.14.2. Series with Nonnegative Terms 428
3.14.3. Alternating Series 434
3.14.4. Absolute and Conditional Convergence of Series 435
3.14.5. Power Series 437
3.14.6. Series of Complex Numbers 449
3.14.7. Fourier Series 457

Chapter 15. Ordinary Differential Equations 465

3.15.1. First-Order Differential Equations 465
3.15.2. Differential Equations of the Second and Higher Orders 480
3.15.3. Second-Order Linear Differential Equations 485
3.15.4. Second-Order Linear Differential Equations with Constant Coefficients 493
3.15.5. Employing Linear Differential Equations When Studying Oscillatory Motions 501

Index 507

No human investigation can be called
a true science if it was not
proved by mathematical means.

*Leonardo da Vinci*

## INTRODUCTION

Mathematics is the most ancient and at the same time the youngest science. It originated in the second century B.C. when the needs of trade, land surveying and navigation made it necessary to put in order the techniques of computation and measurement, techniques whose origin can be traced to still more ancient times. Even the builders of the pyramids in Egypt had some mathematical knowledge.

In Ancient Greece, mathematics received the status of independent science beginning with the sixth century B.C. In the third century, in his immortal *Elements*, Euclid conclusively legalized mathematics as a science. For more than two thousand years this book or its more intelligible expositions were used to study geometry. It can be seen that mathematics differs from all other sciences. Although Aristotle's views on physics seem somewhat naive nowadays and have passed into history, they generalized the knowledge of the surrounding world available at that time, and the Pythagorean theorem still lies at the basis of geometry.

Having been put into shape, mathematics continued to develop. New methods were elaborated, new fields discovered, the symbols and scientific apparatus were perfected. The origination of the physics of New Times was closely associated with the use of mathematics by Kepler and Galileo in their studies of celestial and terrestrial phenomena. The great turning point in the history of mathematics was in the seventeenth century when Descartes created analytic geometry, and Newton and Leibniz, differential and integral calculus. To a large degree, these discoveries opened up possibilities for the development of both mathematics and other sciences such as physics and astronomy.

The rapid subsequent development of mathematics led, at the turn of this century, to a scientific revolution connected, in particular, with the recognition of the validity of non-Euclidean geometries (those of Lobachevsky, Riemann and Bolyai) and with the creation of the set theory by Cantor. Even now mathematics continues to develop, striking our imagination by the variety of special fields,

the newness and extraordinarity of the representations and concepts used, by the unexpected originality of methods and the peculiarity of the language. The evolution and differentiation of sciences involved mathematics too, giving rise to many branches within it.

Alongside the development of the methods and branches of mathematics, the process of its penetration into other sciences, the so-called mathematization of science, has gone on. By virtue of the logic of the development of the science itself, mathematics turned into a method of research. Whereas in the period of classical physics, mathematics mainly served as an instrument for processing experimental data and establishing an exact quantitative relations between physical phenomena and processes, by the end of the nineteenth century, mathematical computations already forestalled hypotheses and discoveries of physics. Mathematics was used not only to process the readings of instruments and the results of experiments but also to produce models whose actual physical meaning was not yet known and was to be revealed.

This fact was often incorrectly interpreted by scientists themselves as well as by the idealistic philosophers and the well-known aphorism "matter disappears, only equations remain" was coined. Its true meaning and distortions were revealed by Lenin in his classical paper *Materialism and Empirio-Criticism.*

The essence of this phenomenon is that using mathematical methods, we can penetrate the fields of physical world inaccessible for physical methods, discover mathematical regularities in them, produce mathematical models of the unknown physical processes and thus direct the thought of the experimenter. At present a theoretical physicist is, first of all, a mathematician. For a physicist, mathematics is not only an instrument which can be used to describe any phenomenon quantitatively but also a basic source of ideas and principles which generate new theories. Thus, for instance, such theoretical physicists of world renown as Landau and Keldysh were outstanding mathematicians.

Einstein's general theory of relativity is a striking example of the part played by mathematical thought in the discoveries of physics. His theory preceded empirical verification and, when observations were made, they practically coincided with the predictions of the theory.

A no less convincing example is the history of quantum mechanics, which was first constructed by purely mathematical means on the basis of some known but unexplained physical data. It was not only substantiated by appropriate experiments but proved to be a source and stimulus for the further development of the physics of microcosm with all its striking results, whose role became well known in the twentieth century.

All those who studied the history of mathematics and its appli-

cations in the natural sciences and technology and who pondered over the relations between mathematics and the objective world, whether they were mathematicians themselves, physicists or philosophers, invariably came across the wonderful capacity of mathematics to give a correct description or representation of physical processes and of the behaviour of physical universe. Scientists who stood at the cradle of learning, such as Kepler and Galileo, were struck by the results they had achieved and believed that the book of nature had been written by its divine creator in the language of mathematics so that it remained for the scientist only to read the notes. Since that time many propositions have been made concerning the nature of mathematics and its cognitive capacity, but none of them has been generally recognized.

In the twentieth century, a Western philosopher, Wittgenstein, advanced a striking doctrine that mathematics was nothing but a collection of tautologies and mathematical proofs were tautologic transformations. This theory explained the absolute authenticity of mathematics and its versatile applicability. But it was powerless to explain the capacity of mathematics to discover new phenomena in the world, i.e. its capacity which is the most important for the development of science and which makes it possible to use mathematics in the sciences more and more often.

It should be pointed out that mathematics manipulates not only abstractions (like all sciences) but abstractions of a rather high degree. Even an ordinary natural number, say, 4, is an abstraction which distracts attention from all the peculiar features of any four objects (trees, legs of a table, corners of a house, etc.) and only characterizes a class which has four members. As to the concept of a natural number, this is a still higher abstraction since it is a class of all classes which have no less than one member.

The power of mathematics is its capacity to create higher and higher abstractions, to operate with them and study their properties and laws. It is this that makes mathematics so useful in sciences other than physics, as those sciences themselves become theoretical, i.e. begin to create sufficiently high abstractions and make use of them.

The eighteenth-century German philosopher Immanuel Kant said that the more mathematics a science contained the more it deserved to be called a science. At that time mathematics was an inalienable element of only mechanics, physics, and astronomy. The theoretical level of sciences has now become so high and the methods of mathematics so versatile and perfect that their merging is not only possible but even absolutely necessary both for the development of those sciences and for mathematics itself.

Naturally, the process of mathematization has affected different sciences to different degrees. Mathematical methods are used in the inorganic natural sciences and in biological research. This has be-

come possible mainly due to the penetration of biology into the intracellular processes and their analysis on the molecular level. We can cite as an example the investigation of the functioning and the construction of models of some functions of a neuron and the study of the problems of heredity and that of deciphering the genetic code.

In social sciences, which were the most isolated from mathematics except for the application of statistical methods in the study of some social processes and phenomena, we can also cite fields such as demography and linguistics where the application of mathematics has produced good results. But it seems to us that the most significant scientific achievement was the penetration of mathematical methods into economics and the control of economic processes. At present, the scientific control of these processes is only possible with the application of exact mathematical methods in every sphere of the economy, from mineral deposit forecasting to studying the demand for goods and services, from investigating the demand for manpower to planning transport arteries, passenger traffic and experiments concerning the influence exerted by man on atmospheric phenomena. In short, the modern life is now unthinkable without mathematics.

Which mathematics do we mean here, however, the "pure" or "applied" mathematics? This traditional differentiation has become more and more conventional and lost its original meaning. It turns out that even the most abstract divisions of pure mathematics may prove to be applicable in very unexpected fields of science and technology. At the same time, the need to solve specific theoretical and practical problems stimulates the elaboration of newer abstract methods and branches of the science of mathematics.

The last few decades have been marked by a rapid development of the means and methods of computation. With the aid of the methods of computational experiments, mathematical modelling makes it possible to calculate processes which are inaccessible even for experiments (e.g. controlled thermonuclear fusion, plasma physics, laser physics). "Actually a new trend has taken shape in theoretical physics in the last two decades," Academician A.A. Samarsky states. "On the basis of a mathematical model, and using computers, mechanisms and physical processes are investigated, their behaviour is 'played through' under different conditions, optimal parameters and modes of operation of structures already functioning or those being designed are found out. Mathematical prediction is now possible for complicated phenomena and mechanisms, a difficult task with the use of other means."* Completely new possibilities have been opened up for mathematics.

---

* Samarsky, A.A. "Mathematical Modelling and Computational Experiment". *Vestnik Akad. Nauk SSSR*, 5, 1979.

The epoch of the scientific and technical revolution is the epoch of mathematization of science, technology, economics and management, and this defines the place of mathematics in the system of higher education. A present-day researcher or engineer must not only know the fundamentals of mathematics but also have a good command of all the modern mathematical methods of investigation which can be used in the field of his activity. No serious research or engineering work is possible nowadays without mathematics. Karl Marx once said that science achieved perfection only when it began to use mathematics. We can safely say that the study of mathematics promotes the formation of present-day scientific thought, and its wide use is the condition for progress in science and technology.

# Part One

# ANALYSIS OF FUNCTIONS OF ONE VARIABLE

## Chapter 1

## REAL NUMBERS

### 1.1.1. Sets. Notation. Logical Symbols

The concept of a *set* is one of the fundamental concepts in mathematics. It belongs to the so-called primary, indefinable concepts. The words "collection", "family", "system" are synonyms of the word "set". A set of students in a lecture-hall, a collection of students who received only excellent and good marks at the entrance examinations, the family of stars that form the Great Bear, a system of three equations with three unknowns, the set of all integers etc. are all examples of sets. It follows from these examples that a set can contain a finite or infinite number of arbitrary objects.

The objects a set consists of are called its *elements*, or *points*. Sets are usually denoted by capital letters and their elements by small letters. If $x$ is an element of the set $X$, then we write $x \in X$ ($x$ belongs to $X$). If $x$ is not an element of the set $X$, then we write $x \notin X$ ($x$ does not belong to $X$). If $x_1, \ldots, x_n$ are some elements, then the notation $X = \{x_1, \ldots, x_n\}$ means that the set $X$ consists of the elements $x_1, \ldots, x_n$. The notation $X = \{x_1, x_2, x_3, \ldots\}$ has a similar meaning.

Let $X$ and $Y$ be two sets. If $X$ and $Y$ consist of the same elements, then we say that they coincide and write $X = Y$. If $X$ does not contain any elements which do not belong to $Y$, then we say that $X$ is contained in $Y$, or that $X$ is a subset of the set $Y$. In this case we write $X \subset Y$ or $Y \supset X$ ($Y$ contains $X$). If $X$ is not contained in $Y$, then we write $X \not\subset Y$. Mathematics often manipulates empty sets. An *empty set* does not contain any elements and is designated as $\phi$. An empty set is a subset of any set.

In what follows, we shall deal with various sets of real numbers. For short, when no ambiguity arises, we shall write numbers meaning real numbers.

Let $P(x)$ be some property of the number $x$. Then the notation $\{x \mid P(x)\}$ is used for the set of all numbers which possess the property $P(x)$. For example, the set $\{x \mid x^2 - 3x + 2 = 0\}$ is the collection of roots of the equation $x^2 - 3x + 2 = 0$, i.e. this set consists of two elements 1 and 2; $\{x \mid 3 < x < 7\}$ is the set of all

numbers which satisfy the inequalities $3 < x < 7$; $\{x \mid x > 7$ and $x < 3\} = \phi$, i.e. this is an empty set.

If $x_1, \ldots, x_n$ are arbitrary numbers, then the notation $x = \max\{x_1, \ldots, x_n\}$ ($x = \min\{x_1, \ldots, x_n\}$) means that the number $x$ is the maximum (minimum) of the numbers $x_1, \ldots, x_n$.

Separate words or whole expressions are often repeated in mathematical propositions (definitions, theorems, etc.) and therefore it is convenient to use an economical logical symbolic notation.

We shall give here only several most simple and frequently used logical symbols. Instead of the words "there exists" or "there is", we use the symbol $\exists$ (the upturned letter $E$ from the word "existence") and instead of the words "any", "every", we use the symbol $\forall$ (the upturned letter $A$ from the word "any"). For instance, the notation $\exists x \in X \ldots$ signifies that "there is a number $x$ from the set $X$ such that . . .". The notation $\forall x \in X$ reads as "for any number $x$ from the set $X$". The notation $\forall x \in X$: $\alpha$ reads as "for any number $x$ from the set $X$ the statement $\alpha$ holds true (is satisfied)".

To make easier the understanding and reading of statements written with the use of logical symbols, we write in parentheses all that refers only to that statement. Thus, for instance, the notation

$$(\forall \varepsilon > 0)\ (\exists \delta > 0)\ (\forall x \neq x_0,\ |x - x_0| < \delta):\quad |f(x) - A| < \varepsilon$$

must be read as follows: "for any $\varepsilon > 0$ there is $\delta > 0$ such that the inequality $|f(x) - A| < \varepsilon$ holds true for all $x$ which are unequal to $x_0$ and satisfy the inequality $|x - x_0| < \delta$".

The symbol ■ used in the text means the end of a proof.

### 1.1.2. Real Numbers and Their Principal Properties

We get a notion of real numbers from the course in elementary mathematics and so we know that the set of real numbers consists of rational and irrational numbers. A *rational number* is a number which can be represented in the form $p/q$, where $p$ and $q$ are integers and $q \neq 0$. An *irrational number* is any real number which is not rational. Every rational number either is an integer or can be represented by a finite or periodic nonterminating decimal fraction. An irrational number can be represented by a nonperiodic nonterminating decimal fraction. For example, the rational numbers 3/4 and 1/3 can be represented as the decimal fractions $3/4 = 0.75$ and $1/3 = 0.333\ldots$; the irrational numbers $\sqrt{2}$ and $\pi$ can be represented as the nonperiodic nonterminating decimal fractions $\sqrt{2} = 1.41421356\ldots$ and $\pi = 3.14159\ldots$ .

We shall systematize the data concerning real numbers, enumerate their principal properties and then obtain some consequences from them.

**I. Addition and multiplication of real numbers.**

For any pair $a$ and $b$ of real numbers, two real numbers $a + b$ and $a \cdot b$ are defined, in a unique way, and are called their *sum* and their *product* respectively, and the following properties hold true. Whatever the real numbers $a$, $b$, $c$,

1°. $a + b = b + a$ (*commutative property*).

2°. $a + (b + c) = (a + b) + c$ (*associative property*).

3°. $a \cdot b = b \cdot a$ (*commutative property*).

4°. $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ (*associative property*).

5°. $(a + b) \cdot c = ac + bc$ (*distributive property*).

6°. *For any number $a$ there is a unique number $0$ such that $a + 0 = a$.*

7°. *For any number $a$ there is a number $(-a)$ such that $a + (-a) = 0$.*

8°. *There is a unique number $1 \neq 0$ such that for any number $a$ the equality $a \cdot 1 = a$ holds true.*

9°. *For any number $a \neq 0$ there is a number $a^{-1}$ such that $a \cdot a^{-1} = 1$; the number $a^{-1}$ can also be designated as $\frac{1}{a}$.*

**II. Comparison of real numbers.**

For any two real numbers $a$ and $b$ one of the following relations is established: $a = b$ ($a$ is equal to $b$), $a > b$ ($a$ is greater than $b$) or $b > a$. The relation "=" possesses the following property: if $a = b$ and $b = c$, then $a = c$.

The relation ">" possesses the following properties: whatever the numbers $a$, $b$, $c$,

10°. *if $a > b$ and $b > c$, then $a > c$;*

11°. *if $a > b$, then $a + c > b + c$;*

12°. *if $a > 0$ and $b > 0$, then $a \cdot b > 0$.*

Instead of $a > b$ we can write $b < a$ ($b$ is smaller than $a$). The notation $a \geqslant b$ (or, what is the same, $b \leqslant a$) signifies that either $a = b$ or $a > b$. The relations $a < b$, $a \leqslant b$, $a > b$, $a \geqslant b$ are called *inequalities*. The inequalities $a < b$ and $a > b$ are *strict inequalities*.

**III. Continuity of real numbers.**

13°. *Let $X$ and $Y$ be two sets consisting of real numbers. If the inequality $x \leqslant y$ is satisfied for any numbers $x \in X$ and $y \in Y$, then there is at least one number $c$ such that the inequalities $x \leqslant c \leqslant y$ are satisfied for all $x$ and $y$ of this kind.*

Note that the set of all real numbers possesses the property of continuity, whereas the set of rational numbers does not possess it. Indeed, assume that the set $X$ consists of rational numbers $x$ for which the inequality $x < \sqrt{2}$ is satisfied and the set $Y$ consists of

rational numbers $y$ for which the inequality $y > \sqrt{2}$ is satisfied. Then, evidently, the inequality $x \leqslant y$ is satisfied for any $x \in X$ and any $y \in Y$, but there is no rational number $c$ such that the inequalities $x \leqslant c \leqslant y$ are satisfied for all $x$ and $y$ of this kind. Indeed, only $\sqrt{2}$ could be this number but, as is known, it is not rational.

All other properties of real numbers follow from Properties I, II, III. We shall get acquainted with only some of them, but in what follows we shall also use other properties without carrying out the formal proof.

Whatever the numbers $a$, $b$, $c$, $d$:

14°. *The number $x = b+(-a)$ is a solution of the equation $a + x = b$.*

Indeed, according to Properties 1°, 2°, 6°, 7°, we have $a + b + (-a) = b$. ■

The number $b + (-a)$ is the *difference* of the numbers $b$ and $a$ and is designated as $b - a$. Note that if $a < b$ (or, what is the same, $b > a$), then the difference $b - a > 0$. Indeed, by virtue of Property 11°, we find from the inequality $b > a$ that $b + (-a) > a + (-a)$ or $b - a > 0$.

15°. *The number $x = ba^{-1}$ is a solution of the equation $ax = b$ if $a \neq 0$.*

Indeed, according to Properties 3°, 4°, 8°, 9°, we have $a \cdot ba^{-1} = b$. ■

The number $ba^{-1}$ is the *quotient* of the division of $b$ by $a$ and is designated as $\frac{b}{a}$ or $b:a$.

16°. *If $a < b$, then $-a > -b$.*

Indeed, since $a < b$, it follows that $b - a > 0$. Consequently, on the basis of Property 11°, $b - a + (-b) > 0 + (-b)$, whence we find that $-a > -b$. ■

In particular, if $a > 0$, then $-a < 0$, and if $a < 0$, then $-a > 0$ (we have used here the fact that $-0 = 0$; indeed, on the basis of Property 6°, $(-0) + 0 = -0$, and on the basis of Property 7°, $(-0) + 0 = 0$, whence it follows that $-0 = 0$).

17°. *If $a > b$ and $c > d$, then $a + c > b + d$.*

Indeed, if $a > b$ and $c > d$, then, by virtue of Property 11°, we have $a + c > b + c$ and $c + b > d + b$. Therefore, on the basis of Property 10°, we have $a + c > b + d$. ■

18°. *If $a < b$ and $c > d$, then $a - c < b - d$.*

Indeed, since $c > d$, it follows, by virtue of Property 16°, that $-c < -d$. Adding the inequalities $a < b$ and $-c < -d$ term by term (we can do this by virtue of Property 17°), we get $a - c < b - d$. ■

19°. $a - a = 0$.

Indeed, $a - a = a + (-a) = 0$. ■

20°. $a \cdot 0 = 0$.

Indeed, $a \cdot 0 = a \cdot (b - b) = ab - ab = 0$. ■

21°. $-(-a) = a$.

Indeed, $-(-a) = (-(-a)) + (-a) + a = 0 + a = a$. ■

22°. $(-a)\, b = -ab$.

Indeed, $(-a)\, b = (-a)\, b + ab + (-ab) = [(-a) + a] \cdot b - ab = 0 \cdot b - ab = 0 - ab = -ab$. ■

Note that when we replaced the sum $(-a)\, b + ab$ by the product $[(-a) + a] \cdot b$, we used Property 5°. From Property 22° we get, in particular, $(-1)\, a = -a$.

23°. *If* $a < 0$ *and* $b > 0$, *then* $ab < 0$.

Indeed, since $a < 0$, it follows that $-a > 0$, and therefore, by virtue of Property 12°, $(-a)\, b > 0$. Consequently, $(-a)\, b = -ab > 0$ and, hence, $ab < 0$. ■

24°. *If* $a < 0$ *and* $b < 0$, *then* $ab > 0$.

Indeed, since $b < 0$, it follows that $-b > 0$, and therefore, by virtue of Property 23°, $(-b)\, a < 0$. Consequently, $(-b)\, a = -ab < 0$ and, hence, $ab > 0$. ■

25°. *If* $a \neq 0$, *then* $a \cdot a = a^2 > 0$.

The validity of this assertion follows from Properties 12° and 24°. In particular, $1 = 1^2 > 0$, i.e. $1 > 0$.

26°. *If* $a > 0$, *then* $a^{-1} > 0$.

Indeed, by virtue of Properties 9° and 25°, we have $aa^{-1} = 1 > 0$, and if we assume that $a^{-1} \leqslant 0$, then, by virtue of Properties 20° and 23°, we have $aa^{-1} \leqslant 0$, i.e. we obtained a contradiction. Consequently, $a^{-1} > 0$. ■

Thus we see that the principal properties I-III of real numbers yield all their other properties. We can therefore say that real numbers constitute a set of elements which possess Properties I-III. This definition of real numbers is known as an *axiomatic property* and Properties I-III as the *axioms* of real numbers.

It should be pointed out in conclusion that, proceeding from Properties I-III, we can represent any real number as a nonterminating decimal fraction. We shall not discuss this problem here, however.

### 1.1.3. Geometric Representation of Real Numbers

**1. Representation of real numbers by points on the coordinate axis.** We shall introduce some preliminary notions. Given an arbitrary straight line, we can indicate two mutually perpendicular directions on it. We choose one of them and denote it by an arrow on the figure (Fig. 1). In addition, we choose a scale unit to measure the lengths of line segments. A straight line with a direction chosen on it is called an *axis*.

We consider two arbitrary points $A$ and $B$ on the axis. A line segment with boundary points $A$ and $B$ is called a *directed segment*

if it is indicated which of the points, $A$ or $B$, is the origin and which is the terminus of the segment. We designate a directed segment with the origin at the point $A$ and the terminus at the point $B$ as $\overrightarrow{AB}$ and assume that it is directed from the origin to the terminus. Note that in the notation $\overrightarrow{AB}$ the letter which denotes the origin of the directed

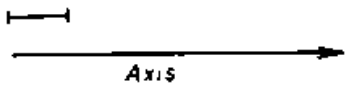



Fig. 1

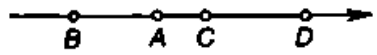



Fig. 2

segment is written *first* and that which denotes the terminus, *second*. The length of the directed segment $\overrightarrow{AB}$ is designated as $|\overrightarrow{AB}|$ or $|AB|$.

For directed segments which lie on the axis (or are parallel to the axis) we introduce the concept of the value of a directed segment. The *value* $AB$ of the directed segment $\overrightarrow{AB}$ is the number which is equal to $|\overrightarrow{AB}|$ if the directions of the segment and of the axis coincide and to $-|\overrightarrow{AB}|$ if the directions are opposite. For the segments $\overrightarrow{AB}$ and $\overrightarrow{CD}$, shown in Fig. 2, $AB = -|\overrightarrow{AB}|$ and $CD = |\overrightarrow{CD}|$.

Note that the values of the directed segments $\overrightarrow{AB}$ and $\overrightarrow{BA}$ differ in sign for any direction of the axis, i.e.

$$AB = -BA.$$

If the points $A$ and $B$ coincide, then we assume the value of the directed segment $\overrightarrow{AB}$ to be equal to zero.

For any three points $A$, $B$, $C$ lying on the axis there holds an equality

$$AB + BC = AC,$$

which is known as the *fundamental identity* (we use it repeatedly in what follows).

We can easily establish the validity of the fundamental identity using the figure, but then we must consider various cases of the mutual positions of the points $A$, $B$, $C$ on the axis. If all the three points $A$, $B$, $C$ are distinct, then there are six cases (Fig. 3). It is easy to verify the fundamental identity in each of these cases.

We pass now to the geometric representation of real numbers. Let us consider a straight line. We choose a direction on it (then it becomes an axis) and a point $O$ (the origin). A straight line with the chosen direction and origin is a *coordinate axis* (we assume a scale

unit to be chosen). Let $M$ be an arbitrary point on the straight line (Fig. 4*a*). We associate the point $M$ with a number $x$ which is equal to the value $OM$ of the directed segment $\overrightarrow{OM}$. The number $x$ is the *coordinate* of the point $M$. Thus every point of the coordinate axis is associated with a definite real number, its coordinate. The converse

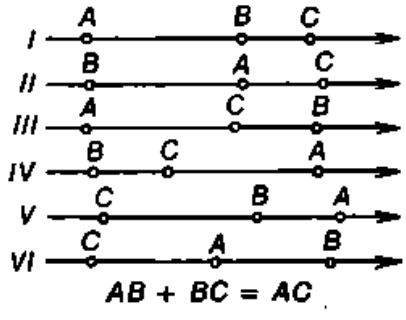



Fig. 3

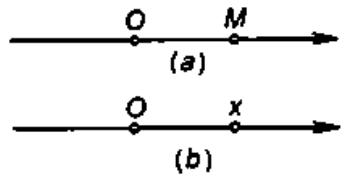



Fig. 4

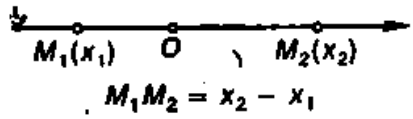



Fig. 5

is also true: to every real number $x$ there corresponds a certain point of the coordinate axis, namely, a point $M$ whose coordinate is equal to $x$.

Thus real numbers can be represented by points on the coordinate axis and therefore a number, i.e. its coordinate, is often written beside a point on the coordinate axis (Fig. 4*b*).

Assume that a point $M_1$ has a coordinate $x_1$ and a point $M_2$, a coordinate $x_2$ (Fig. 5). We express the value $M_1M_2$ of the directed segment $\overrightarrow{M_1M_2}$ in terms of the coordinates of the points $M_1$ and $M_2$. In accordance with the fundamental identity,

$$OM_1 + M_1M_2 = OM_2,$$

whence it follows that $M_1M_2 = OM_2 - OM_1$. But $OM_1 = x_1$, $OM_2 = x_2$, and therefore

$$M_1M_2 = x_2 - x_1.$$

We shall often use this formula in analytic geometry.

**2. Some frequently used number sets.** Let $a$ and $b$ be two numbers and $a < b$. We shall use the following designations:

$$\{x \mid a \leqslant x \leqslant b\} = [a, b], \ \{x \mid a < x \leqslant b\} = (a, b],$$
$$\{x \mid a \leqslant x < b\} = [a, b), \ \{x \mid a < x < b\} = (a, b),$$
$$\{x \mid a \leqslant x\} = [a, +\infty), \ \{x \mid a < x\} = (a, +\infty),$$
$$\{x \mid x \leqslant b\} = (-\infty, b], \ \{x \mid x < b\} = (-\infty, b).$$

We shall designate the set of all real numbers as follows:

$$\{x \mid -\infty < x < +\infty\} \text{ or } (-\infty, +\infty).$$

All these sets are known as *intervals*, of which $[a, b]$ is a *closed interval*, $[a, b)$, $(a, b]$, $[a, +\infty)$, $(-\infty, b]$ are *half-open* or *half-closed intervals*, $(a, b)$, $(a, +\infty)$, $(-\infty, b)$, $(-\infty, +\infty)$ are *open intervals*. The intervals $[a, b]$, $(a, b]$, $[a, b)$, $(a, b)$ are *finite intervals*, $a$ and $b$ being their *endpoints*. The other intervals are *infinite*.

Numerical intervals are associated with intervals on the coordinate axis. For example, the closed interval $[x_1, x_2]$ is represented on the coordinate axis by a line segment $M_1M_2$ such that the point $M_1$ has a coordinate $x_1$ and the point $M_2$, a coordinate $x_2$ (see Fig. 5). The set $(-\infty, +\infty)$ of all numbers is represented by the whole coordinate axis, and therefore the set $(-\infty, +\infty)$ is also known as the *number line* (or *number axis*) and any number is a *point* of that line. Assume that $a$ is an arbitrary point of the number line and $\delta$ is a positive number. The interval $(a - \delta, a + \delta)$ is known as the $\delta$-*neighbourhood of the point* $a$.

### 1.1.4. Bounds of Number Sets

A set $X$ is said to be bounded from above (from below) if there is a number $c$ such that the inequality $x \leqslant c$ $(x \geqslant c)$ is satisfied for any $x \in X$. In this case the number $c$ is the *upper* (*lower*) *bound* of the set $X$.

A set which is bounded both from above and from below is said to be *bounded*. Thus, for instance, any finite interval ($[a, b]$, $[a, b)$, $(a, b]$, $(a, b)$) is bounded. The interval $(a, +\infty)$ is a set bounded from below but not bounded from above, and the entire number line $(-\infty, +\infty)$ is a set not bounded either from above or from below.

Any set $X$ bounded from above (from below) evidently has infinitely many upper (lower) bounds. Indeed, if the number $c$ is an upper (lower) bound of the set $X$, then any number $c'$, which is larger (smaller) than the number $c$, is also an upper (lower) bound of the set $X$ since the validity of the inequality $x \leqslant c$ $(x \geqslant c)$ implies the validity of the inequality $x \leqslant c'$ $(x \geqslant c')$.

A question naturally arises concerning the existence of the least upper bound of a set bounded from above and the greatest lower bound of a set bounded from below.

The least upper bound of a set $X$ bounded from above is called the *supremum* of the set $X$ and is designated as $\sup X$ and the greatest lower bound of a set $X$ bounded from below is called the *infimum* of the set and is designated as $\inf X$.

**Example.** Let $X = (a, b)$. Then the number $b$ is the least upper bound of the set $X$ and the number $a$ is its greatest lower bound,

i.e. $b = \sup X$ and $a = \inf X$. Let now $X = (a, +\infty)$. Then $a = \inf X$ and there are no upper bounds, including the least upper bound, of this set.

The least upper bound ($\sup X$) possesses the following significant property. However small the number $\varepsilon > 0$, there is always $x \in X$ such that $x > \sup X - \varepsilon$*. In fact, if there were no number of this kind, then the number $X - \varepsilon$ would also be an upper bound of the set $X$ and then the number $\sup X$ would not be the least upper bound. In other words, this property expresses the fact that the number $\sup X$ is the least of the numbers which bound the set $X$ from above and cannot be made smaller.

This property of the least upper bound can be reformulated as follows: if $c = \sup X$, then for any number $c' < c$ there is a number $x \in X$ such that $x > c'$**. To make sure of the equivalence of these two formulations, it is sufficient to take $c'$ and $\varepsilon$ connected by the relation $c' = c - \varepsilon$ from which it follows that the condition $\varepsilon > 0$ is equivalent to the condition $c' < c$.

The greatest lower bound possesses a similar property, i.e. however small the number $\varepsilon > 0$, there is a number $x \in X$ such that $x < \inf X + \varepsilon$. (The reader is invited to give the other formulation of this property.)

A question arises as to whether a set bounded from above (from below) always has the least upper (the greatest lower) bound. The following significant theorem answers this question.

**Theorem 1.1.** *Any nonempty number set bounded from above (from below) has the least upper (the greatest lower) bound.*

*Proof.* Let $X$ be a nonempty set bounded from above. Then the set $Y$ of numbers which bound $X$ from above is nonempty. It follows from the definition of an upper bound that for any $x \in X$ and any $y \in Y$ the inequality $x \leqslant y$ holds true. By virtue of the property of the continuity of real numbers there is a number $c$ such that the inequalities

$$x \leqslant c \leqslant y \tag{1}$$

hold true for any $x$ and $y$. It follows from the first of inequalities (1) that the number $c$ bounds the set $X$ from above, i.e. is its upper bound, and it follows from the second inequality that it is the least of these numbers, i.e. the least upper bound.

The case of the existence of the greatest lower bound of a nonempty set bounded from below can be treated by analogy. ■

If the set $X$ is not bounded from above (from below), then we conventionally write $\sup X = +\infty$ ($\inf X = -\infty$).

---

* Or, using logical symbols, $(\forall \varepsilon > 0)\ (\exists x \in X)\colon\ x > \sup X - \varepsilon$.

** Or, using logical symbols, $(\forall c' < c)\ (\exists x \in X)\colon\ x > c'$.

### 1.1.5. The Absolute Value of a Number

The concept of an absolute value of a number and inequalities which are connected with absolute values are often used in what follows.

**Definition.** *The absolute value (or modulus) of a number $x$ is the number $x$ itself if $x \geqslant 0$ or the number $-x$ if $x < 0$.*

The absolute value of the number $x$ is designated as $|x|$. Thus

$$|x| = \begin{cases} x \text{ if } x \geqslant 0, \\ -x \text{ if } x < 0. \end{cases}$$

This definition yields a number of properties of the absolute value of a number.

1°. $|x| \geqslant 0$. Indeed,

(1) if $x \geqslant 0$, then $|x| = x \geqslant 0$;

(2) if $x < 0$, then $|x| = -x$; but $-x > 0$ since $x < 0$, i.e. $|x| > 0$. From (1) and (2) we find that $|x| \geqslant 0$.

2°. $|x| = |-x|$. Indeed,

(1) if $x \geqslant 0$, then $-x \leqslant 0$ and then $|-x| = -(-x) = x = |x|$;

(2) if $x < 0$, then $-x > 0$ and then $|-x| = -x = |x|$ since $x < 0$. From (1) and (2) we find that $|x| = |-x|$.

3°. $-|x| \leqslant x \leqslant |x|$. Indeed,

(1) if $x \geqslant 0$, then $|x| = x$ and $-x \leqslant 0$. Hence $|x| \geqslant -x$, i.e. $-|x| \leqslant x = |x|$;

(2) if $x < 0$, then $|x| = -x$, whence $-|x| = x$. Furthermore, since $x < 0$, we have $2x < 0$, or $x + x < 0$, whence $x < -x$, i.e. $x < |x|$. Thus $-|x| = x < |x|$. From (1) and (2) we find that $-|x| \leqslant x \leqslant |x|$.

Since the three properties that follow are very significant, we shall prove them as theorems.

**Theorem 1.2.** *Let $\varepsilon$ be a positive number. Then the inequalities $|x| \leqslant \varepsilon$ and $-\varepsilon \leqslant x \leqslant \varepsilon$ are equivalent.**

*Proof.* Let $|x| \leqslant \varepsilon$. Then

(1) if $x \geqslant 0$, then $|x| = x$ and, hence, $x \leqslant \varepsilon$, whence $0 \leqslant x \leqslant \varepsilon$;

(2) if $x < 0$, then $|x| = -x$, and, hence, $-x \leqslant \varepsilon$, whence $-\varepsilon \leqslant x \leqslant 0$. Combining (1) and (2), we get $-\varepsilon \leqslant x \leqslant \varepsilon$ for any $x$.

Assume that the inequalities $-\varepsilon \leqslant x \leqslant \varepsilon$ hold true. This means that the inequalities $x \leqslant \varepsilon$ and $x \geqslant -\varepsilon$ simultaneously hold true. From the last inequality we have $-x \leqslant \varepsilon$. Since $|x|$ is either $x$ or $-x$ by definition, it follows that $|x| \leqslant \varepsilon$. ■

---

* Using logical symbols, we can write these statements as follows: $\forall \varepsilon > 0$ $|x| \leqslant \varepsilon \Leftrightarrow -\varepsilon \leqslant x \leqslant \varepsilon$.

**Theorem 1.3.** *The absolute value of the sum of two numbers is not larger than the sum of the absolute values of those numbers, i.e.* $|x+y| \leqslant |x| + |y|$.

*Proof.* Let $x$ and $y$ be some numbers. According to Property 3°, the inequalities

$$-|x| \leqslant x \leqslant |x|, \quad -|y| \leqslant y \leqslant |y|$$

hold true for them. Summing them up term by term, we obtain

$$-(|x| + |y|) \leqslant x + y \leqslant (|x| + |y|).$$

In accordance with Theorem 1.2, this two-sided inequality is equivalent to the inequality $|x+y| \leqslant |x| + |y|$. ■

Note that $|x-y| \leqslant |x| + |y|$.

**Theorem 1.4.** *The absolute value of the difference of two numbers is not smaller than the difference of the absolute values of the numbers, i.e.* $|x-y| \geqslant |x| - |y|$.

*Proof.* For any numbers $x$ and $y$ we have

$$x = y + (x - y).$$

In accordance with Theorem 1.3, there holds an inequality

$$|x| = |y + (x-y)| \leqslant |y| + |x-y|,$$

whence we obtain $|x-y| \geqslant |x| - |y|$. ■

Note that $|x+y| \geqslant |x| - |y|$.

It should be pointed out in conclusion that whatever the two numbers $x$ and $y$, there hold relations which are easy to verify:

$$|x \cdot y| = |x| \cdot |y|, \quad \left|\frac{x}{y}\right| = \frac{|x|}{|y|} \text{ if } y \neq 0.$$

## Chapter 2

## THE LIMIT OF A SEQUENCE

The concept of a limit and that of a function are fundamental concepts of analysis. We begin to study limits in elementary mathematics where, with the aid of limiting processes, we determine the length of a circle, the volume of a cylinder or a cone etc. It is also used when we define the sum of an infinitely decreasing geometric progression. The limiting process is one of the basic operations of analysis. In this chapter we shall consider the simplest form of the limiting processes based on the concept of the limit of a number sequence. The latter concept will make it possible to define other, more complicated, limiting processes.

### 1.2.1. Number Sequences

**1. Number sequences and arithmetic operations involving them.** We study number sequences as early as in the secondary school. We can give the following examples of sequences: (1) the sequence of all terms of an arithmetic or geometric progression; (2) the sequence of perimeters of regular $n$-gons inscribed in a given circle; (3) the sequence $x_1 = 1$, $x_2 = 1.4$, $x_3 = 1.41\ldots$ of approximations of $\sqrt{2}$.

Let us elaborate and extend the concept of a number sequence.

**Definition.** *If every number $n$ from the series of natural numbers*

$$1,\ 2,\ 3,\ \ldots,\ n,\ \ldots$$

*is associated with a real number $x_n$, then the set of real numbers*

$$x_1,\ x_2,\ x_3,\ \ldots,\ x_n,\ \ldots \tag{1}$$

*is a number sequence or simply a sequence.**

The numbers $x_1, x_2, x_3, \ldots, x_n, \ldots$ are the *elements* (or *terms*) of sequence (1), the symbol $x_n$ is the *general element* (or *term*) of the

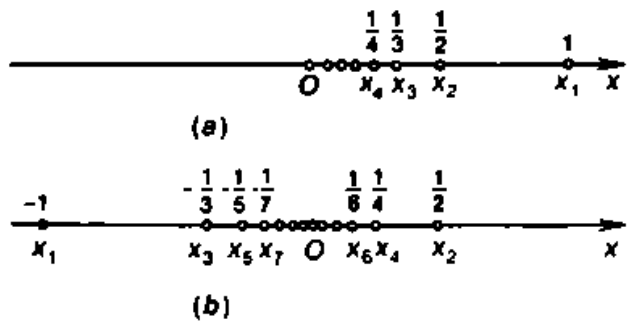



Fig. 6

sequence and $n$ is its *number*. In an abbreviated form sequence (1) is designated as $\{x_n\}$. Thus, for instance, the symbol $\{1/n\}$ denotes the sequence $1,\ 1/2,\ 1/3,\ \ldots,\ 1/n,\ \ldots$.

A sequence is considered to be defined if a way of obtaining any one of its elements is indicated. For example, the formula $x_n = 1 + (-1)^n$ defines the sequence $0.2,\ 0.2,\ \ldots$. Turning the fraction 1/3 into a decimal fraction and leaving one, two, three etc decimal places, we get a sequence

$$x_1 = 0.3,\ x_2 = 0.33,\ x_3 = 0.333,\ \ldots,\ x_n = 0.333\ldots3,\ \ldots.$$

According to the definition, a sequence contains an infinite number of elements, any two of its elements differing at least by their numbers.

* In other words, we can define a number sequence as the set of pairs of numbers $(n,\ x_n)$ in which the first number successively assumes the values $1,\ 2,\ 3,\ \ldots$.

In terms of geometry, a sequence is depicted on the coordinate axis as a sequence of points whose coordinates are equal to the corresponding elements of the sequence. Figure 6*a* and *b* shows the sequences $\{x_n\} = \{1/n\}$ and $\{x_n\} = \{(-1)^n/n\}$ respectively.

We shall discuss now arithmetic operations involving number sequences. Let us consider the sequences $\{x_n\}$ and $\{y_n\}$.

The *product* of the sequence $\{x_n\}$ by the number $m$ is the sequence $mx_1, mx_2, \ldots, mx_n, \ldots$;

the *sum* of the given sequences is the sequence $x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n, \ldots$;

the *difference* is the sequence $x_1 - y_1, x_2 - y_2, \ldots, x_n - y_n, \ldots$;

the *product* is the sequence $x_1y_1, x_2y_2, \ldots, x_ny_n, \ldots$;

the *quotient* is the sequence $\frac{x_1}{y_1}, \frac{x_2}{y_2}, \ldots, \frac{x_n}{y_n}, \ldots$ if all terms of the sequence $\{y_n\}$ are nonzero.

We can write all actions on sequences in the following symbolic form:

$$m\{x_n\} = \{mx_n\},\ \{x_n\} + \{y_n\} = \{x_n + y_n\},$$
$$\{x_n\} - \{y_n\} = \{x_n - y_n\},\ \{x_n\}\cdot\{y_n\} = \{x_ny_n\},$$
$$\frac{\{x_n\}}{\{y_n\}} = \left\{\frac{x_n}{y_n}\right\},\quad y_n \neq 0^*.$$

**2. Bounded and unbounded sequences.**

**Definition 1.** *The sequence $\{x_n\}$ is said to be bounded from above (from below) if there is a number $M$ (a number $m$) such that any element $x_n$ of this sequence satisfies the inequality $x_n \leqslant M$ ($x_n \geqslant m$).*

**Definition 2.** *The sequence $\{x_n\}$ is said to be bounded if it is bounded both from above and from below, i.e. there are numbers $m$ and $M$ such that any element $x_n$ of this sequence satisfies the inequalities $m \leqslant x_n \leqslant M$.*

Let $A = \max\{|m|, |M|\}$. Then we can write the condition for the boundedness of the sequence as $|x_n| \leqslant A$.

**Definition 3.** *The sequence $\{x_n\}$ is said to be unbounded if for any positive number $A$ there is an element $x_n$ of this sequence which satisfies the inequality $|x_n| > A$ (i.e. either $x_n > A$ or $x_n < -A$).*

It follows from these definitions that if a sequence is bounded from above, then all its elements belong to the interval $(-\infty, M]$; if it is bounded from below, then they belong to the interval $[m, +\infty)$, and if it is bounded both from above and from below, then they belong to the interval $[m, M]$. An unbounded sequence can be bounded from above (from below).

Let us consider some examples of bounded and unbounded sequences.

1. The sequence $1, 2, 3, \ldots, n, \ldots$ is bounded from below but unbounded from above.

* $y_n \neq 0$ means that the values $y_n$ are nonzero for any $n$.

2. The sequence $-1, -2, -3, \ldots, -n, \ldots$ is bounded from above but unbounded from below.

3. The sequence $1, 1/2, 1/3, \ldots, 1/n, \ldots$ is bounded since any element $x_n$ of this sequence satisfies the inequalities $0 \leqslant x_n \leqslant 1$ ($m = 0$, $M = 1$).

4. The sequence $-1, 2, -3, 4, -5, \ldots, (-1)^n n, \ldots$ is unbounded. Indeed, whatever the number $A$, there are always elements, among the elements $x_n$ of this sequence, for which the inequality $|x_n| > A$ is satisfied.

Using logical symbols, we can write these definitions as follows:

the sequence $\{x_n\}$ is bounded from above if $(\exists M)\ (\forall x_n)$: $x_n \leqslant M$;

the sequence $\{x_n\}$ is bounded from below if $(\exists m)\ (\forall x_n)$: $x_n \geqslant m$;

the sequence $\{x_n\}$ is bounded if $(\exists A > 0)\ (\forall x_n)$: $|x_n| \leqslant A$;

the sequence $\{x_n\}$ is unbounded if $(\forall A > 0)\ (\exists x_n)$: $|x_n| > A$.

Comparing the notations for the last two definitions with the use of logical symbols, we see that when negations are constructed, the symbols $\exists$ and $\forall$ replace each other.

**3. Infinitely large and infinitely small sequences.**

**Definition 1.** *The sequence $\{x_n\}$ is infinitely large if for any positive number $A$ there is a number $N$ such that the inequality $|x_n| > A$ is satisfied for $n > N$*.*

The symbolic notation of the definition of an infinitely large sequence is

$$(\forall A > 0)\ (\exists N)\ (\forall n > N)\colon\ |x_n| > A.$$

**Remark.** It is evident that any infinitely large sequence is unbounded. However, an unbounded sequence is not obligatory infinitely large. For instance, the unbounded sequence $1, 2, 1, 3, \ldots, 1, n, 1, n+1, \ldots$ is not infinitely large since for $A > 1$ the inequality $|x_n| > A$ is satisfied not for all elements $x_n$ with odd numbers.

**Definition 2.** *The sequence $\{\alpha_n\}$ is infinitely small (infinitesimal) if for any positive number $\varepsilon$ there is a number $N$ such that the inequality $|\alpha_n| < \varepsilon$ is satisfied for $n > N$.*

The symbolic notation of the definition of an infinitely small sequence is

$$(\forall \varepsilon > 0)\ (\exists N)\ (\forall n > N)\colon\ |\alpha_n| < \varepsilon.$$

**Example 1.** Using Definition 1, we shall prove that the sequence $\{n\}$ is infinitely large.

We take any number $A > 0$. From the inequality $|x_n| = |n| > A$ we find that $n > A$. If we take $N \geqslant A$, then the inequality $|x_n| > A$ will be satisfied for all $n > N$, i.e., in accordance with Definition 1, the sequence $\{n\}$ is infinitely large.

* "For $n > N$" means "for all elements of the sequence with the numbers $n > N$".

**Example 2.** Using Definition 2, we shall prove that the sequence $\{1/n\}$ is infinitely small.

We take any number $\varepsilon > 0$. From the inequality $|\alpha_n| = |1/n| < \varepsilon$ we find that $n > 1/\varepsilon$. If we take $N = [1/\varepsilon]$*, then the inequality $n \geqslant [1/\varepsilon] + 1 > 1/\varepsilon$ will be satisfied for all $n > N$, whence it follows that $1/n = |\alpha_n| < \varepsilon$. Thus, in accordance with Definition 2, the sequence $\{1/n\}$ is infinitely small.

We shall prove the theorem which establishes a relationship between infinitely large and infinitely small sequences.

**Theorem 2.1.** *If $\{x_n\}$ is an infinitely large sequence and all its terms are different from zero, then the sequence $\{1/x_n\}$ is infinitely small and, conversely, if $\{\alpha_n\}$ is an infinitely small sequence and $\alpha_n \neq 0$, then the sequence $\{1/\alpha_n\}$ is infinitely large.*

*Proof.* Let $\{x_n\}$ be an infinitely large sequence. We take any $\varepsilon > 0$ and set $A = 1/\varepsilon$. In accordance with Definition 1, for this $A$ there is a number $N$ such that $|x_n| > A$ for $n > N$. From this we get $|1/x_n| = 1/|x_n| < 1/A = \varepsilon$ for all $n > N$. And this means that the sequence $\{1/x_n\}$ is infinitely small.

The proof of the second part of the theorem can be carried out by analogy. ■

**4. Principal properties of infinitely small sequences.**

**Theorem 2.2.** *The sum and the difference of two infinitely small sequences are infinitely small sequences.*

*Proof.* Let $\{\alpha_n\}$ and $\{\beta_n\}$ be infinitely small sequences. We have to prove that the sequence $\{\alpha_n \pm \beta_n\}$ is infinitely small. Assume that $\varepsilon$ is an arbitrary positive number, $N_1$ is a number beginning with which $|\alpha_n| < \varepsilon/2$, and $N_2$ is a number beginning with which $|\beta_n| < \varepsilon/2$. (According to the definition of an infinitely small sequence, such numbers $N_1$ and $N_2$ do exist.) We take $N = \max\{N_1, N_2\}$; then, for $n > N$ the inequalities $|\alpha_n| < \varepsilon/2$ and $|\beta_n| < \varepsilon/2$ will be simultaneously satisfied. Consequently, for $n > N$ we have

$$|\alpha_n \pm \beta_n| \leqslant |\alpha_n| + |\beta_n| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

This means that the sequence $\{\alpha_n \pm \beta_n\}$ is infinitely small. ■

**Corollary.** *The algebraic sum of any finite number of infinitely small sequences is an infinitely small sequence.*

**Theorem 2.3.** *The product of two infinitely small sequences is an infinitely small sequence.*

*Proof.* Assume that $\{\alpha_n\}$ and $\{\beta_n\}$ are infinitely small sequences. We have to prove that the sequence $\{\alpha_n \cdot \beta_n\}$ is infinitely small. Since the sequence $\{\alpha_n\}$ is infinitely small, for any $\varepsilon > 0$ there is

* The symbol $[x]$ denotes the integral part of the number $x$, i.e. the largest integer that does not exceed $x$. For example, $[1] = 1$, $[3, 1] = 3$, $[0, 7] = 0$, $[-0, 5] = -1$, $[-172.9] = -173$ etc. It is evident that $[x] + 1 > x$.

a number $N_1$ such that $|\alpha_n| < \varepsilon$ for $n > N_1$, and since $\{\beta_n\}$ is also an infinitely small sequence, for $\varepsilon = 1$ there is a number $N_2$ such that $|\beta_n| < 1$ for $n > N_2$. We take $N = \max\{N_1, N_2\}$; then, for $n > N$ both inequalities are satisfied. Consequently, for $n > N$ we have

$$|\alpha_n \cdot \beta_n| = |\alpha_n| \cdot |\beta_n| < \varepsilon \cdot 1 = \varepsilon.$$

This means that the sequence $\{\alpha_n \cdot \beta_n\}$ is infinitely small.

**Corollary.** *The product of any finite number of infinitely small sequences is an infinitely small sequence.*

**Remark.** The quotient of two infinitely small sequences may not be an infinitely small sequence and may not even have sense. For instance, if $\alpha_n = 1/n$ and $\beta_n = 1/n$, then all elements $\{\alpha_n/\beta_n\}$ are equal to unity and the sequence is bounded. If $\alpha_n = 1/n$ and $\beta_n = 1/n^2$, then the sequence $\{\alpha_n/\beta_n\}$ is infinitely large, and if $\alpha_n = 1/n^2$ and $\beta_n = 1/n$, then the sequence is infinitely small. If, beginning with a certain element, the elements of $\{\beta_n\}$ are equal to zero, then $\{\alpha_n/\beta_n\}$ has no sense.

**Theorem 2.4.** *The product of a bounded sequence by an infinitely small sequence is an infinitely small sequence.*

*Proof.* Assume that $\{x_n\}$ is a bounded sequence and $\{\alpha_n\}$ is an infinitely small sequence. We have to prove that the sequence $\{x_n \cdot \alpha_n\}$ is infinitely small. Since the sequence $\{x_n\}$ is bounded, there is a number $A > 0$ such that any element $x_n$ satisfies the inequality $|x_n| \leqslant A$. We take any $\varepsilon > 0$. Since the sequence $\{\alpha_n\}$ is infinitely small, for the positive number $\varepsilon/A$ there is a number $N$ such that the inequality $|\alpha_n| < \varepsilon/A$ is satisfied for $n > N$. Consequently, for $n > N$ we have

$$|x_n \cdot \alpha_n| = |x_n| \cdot |\alpha_n| < A \cdot \frac{\varepsilon}{A} = \varepsilon.$$

This means that the sequence $\{x_n \cdot \alpha_n\}$ is infinitely small. ■

**Corollary.** *The product of an infinitely small sequence by a number is an infinitely small sequence.*

Let us discuss now the concept of the limit of a number sequence, which is one of the most important concepts in analysis.

### 1.2.2. Convergent Sequences

**1. The concept of a convergent sequence.**

**Definition.** *The number $a$ is the limit of the sequence $\{x_n\}$ if for any positive number $\varepsilon$ there is a number $N$ such that the inequality*

$$|x_n - a| < \varepsilon \tag{1}$$

*holds true for $n > N$.*

By logical symbols we can write this definition as follows:

$$(\forall \varepsilon > 0)\ (\exists N)\ (\forall n > N):\ |x_n - a| < \varepsilon.$$

A sequence which has a limit is said to be *convergent*.

If the sequence $\{x_n\}$ converges and has a number $a$ as its limit, then the symbolic notation is

$$\lim_{n\to\infty} x_n = a, \quad \text{or} \quad x_n \to a \quad \text{as} \quad n \to \infty.$$

A sequence which is not convergent is *divergent*.

**Example.** Using the definition of the limit of a sequence, we shall prove that $\lim\limits_{n\to\infty} \frac{n}{n+1} = 1$.

We take an arbitrary number $\varepsilon > 0$. Since $|x_n - 1| = \left|\frac{n}{n+1} - 1\right| = \frac{1}{n+1}$, it follows that to find the values of $n$ which satisfy the inequality $|x_n - 1| < \varepsilon$, it is sufficient to solve the inequality $1/(n+1) < \varepsilon$, whence we find that $n > (1-\varepsilon)/\varepsilon$. Consequently, we can take the integral part of the number $(1-\varepsilon)/\varepsilon$ as $N$, i.e. $N = [(1-\varepsilon)/\varepsilon]$. Then the inequality $|x_n - 1| < \varepsilon$ will be satisfied for all $n > N$. We have thus proved that $\lim\limits_{n\to\infty} \frac{n}{n+1} = 1$.

**Remark 1.** Let the sequence $\{x_n\}$ have a number $a$ as its limit. Then $\{\alpha_n\} = \{x_n - a\}$ is an infinitely small sequence since for any $\varepsilon > 0$ there is a number $N$ such that the inequality $|\alpha_n| = |x_n - a| < \varepsilon$ is satisfied for $n > N$. Consequently, any element $x_n$ of the sequence which has the number $a$ as its limit can be represented in the form

$$x_n = a + \alpha_n, \qquad (2)$$

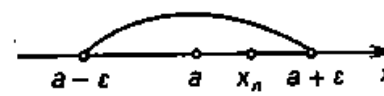



Fig. 7

where $\alpha_n$ is an element of the infinitely small sequence $\{\alpha_n\}$. The converse is evidently also true: if $x_n$ can be written in the form $x_n = a + \alpha_n$, where $\{\alpha_n\}$ is an infinitely small sequence, then $\lim\limits_{n\to\infty} x_n = a$. We can use representation (2) when we prove the theorems on the limits of sequences.

**Remark 2.** Inequality (1) is equivalent to the inequalities

$$-\varepsilon < x_n - a < \varepsilon, \quad \text{or} \quad a - \varepsilon < x_n < a + \varepsilon,$$

which mean that the element $x_n$ is in the $\varepsilon$-neighbourhood of the point $a$ (Fig. 7). Therefore we can formulate the definition of the limit of a sequence as follows: the number $a$ is said to be the limit of the sequence $\{x_n\}$ if for any $\varepsilon$-neighbourhood of the point $a$ there is a number $N$ such that all elements $x_n$ with the numbers $n > N$ are in this $\varepsilon$-neighbourhood.

**Remark 3.** It is evident that the infinitely large sequence $\{x_n\}$ has no limit. We sometimes say that it has a *limit infinity* and

write

$$\lim_{n\to\infty} x_n = \infty.$$

If, in this case, all the terms of the sequence, beginning with a certain term, are positive (negative), then we write

$$\lim_{n\to\infty} x_n = +\infty \quad (\lim_{n\to\infty} x_n = -\infty).$$

As distinct from an infinite limit, we shall call a limit of a sequence in the definition given above a *finite limit*.

**Remark 4.** Every infinitely small sequence is evidently convergent and has the number $a = 0$ as its limit.

**2. Principal properties of convergent sequences.**

Let us prove a lemma which we shall need when proving Theorem 2.5.

**Lemma 2.1.** *If all elements of the infinitely small sequence $\{\alpha_n\}$ are equal to the same number $c$, then $c = 0$.*

*Proof.* Let us assume the contrary, i.e. that $c \neq 0$. We set $\varepsilon = |c|/2$. Then, according to the definition of an infinitely small sequence, there is a number $N$ such that the inequality $|\alpha_n| < \varepsilon$ is satisfied for $n > N$. Since $\alpha_n = c$ and $\varepsilon = |c|/2$, we can rewrite the last inequality in the form $|c| < |c|/2$, whence we find that $1 < 1/2$. The contradiction obtained proves that the inequality $c \neq 0$ cannot hold true and, hence, $c = 0$. ■

**Theorem 2.5.** *A convergent sequence has only one limit.*

*Proof.* Let us assume the contrary, i.e. that the convergent sequence $\{x_n\}$ has two limits $a$ and $b$. Then, from formula (2) for the elements $x_n$, we get

$$x_n = a + \alpha_n, \quad x_n = b + \beta_n,$$

where $\alpha_n$ and $\beta_n$ are elements of the infinitely small sequences $\{\alpha_n\}$ and $\{\beta_n\}$. Making the right-hand sides of these relations equal, we find that $\alpha_n - \beta_n = b - a$. Since all elements of the infinitely small sequence $\{\alpha_n - \beta_n\}$ are equal to the same number $b - a$, it follows, according to Lemma 2.1, that $b - a = 0$, i.e. $b = a$. ■

**Theorem 2.6.** *A convergent sequence is bounded.*

*Proof.* Assume that $\{x_n\}$ is a convergent sequence and the number $a$ is its limit. Assume, furthermore, that $\varepsilon$ is an arbitrary positive number and $N$ is the number beginning with which the inequality $|x_n - a| < \varepsilon$ is satisfied. Then

$$|x_n| = |(x_n - a) + a| \leqslant |x_n - a| + |a| < |a| + \varepsilon$$

for all $n > N$. Let $A = \max\{|a| + \varepsilon, |x_1|, |x_2|, \ldots, |x_N|\}$. It is evident that $|x_n| \leqslant A$ for all numbers $n$, and this means that the sequence $\{x_n\}$ is bounded. ■

**Remark.** A bounded sequence may not be convergent. For instance, the sequence $-1, 1, -1, \ldots, (-1^n), \ldots$ is evidently bounded but does not converge. Let us prove this. We assume that this sequence has a number $a$ as its limit. Then, for $\varepsilon = 1/2$ there is a number $N$ such that $|x_n - a| < 1/2$ for $n > N$. Since $x_n$ alternatively assumes the values 1 and $-1$, it follows that $|1 - a| < 1/2$ and $|(-1) - a| < 1/2$. Using these inequalities, we find that

$$2 = |1 - a + a - (-1)| \leqslant |1 - a| + |a - (-1)| < \frac{1}{2} + \frac{1}{2} = 1,$$

i.e. $2 < 1$. The contradiction obtained proves that the sequence is divergent.

**Theorem 2.7.** *The sum (difference) of two convergent sequences $\{x_n\}$ and $\{y_n\}$ is a convergent sequence whose limit is equal to the sum (difference) of the limits of the sequences $\{x_n\}$ and $\{y_n\}$.*

*Proof.* Let $a$ and $b$ be the limits of the sequences $\{x_n\}$ and $\{y_n\}$ respectively. Then, from formula (2), we find that

$$x_n = a + \alpha_n, \quad y_n = b + \beta_n,$$

where $\{\alpha_n\}$ and $\{\beta_n\}$ are infinitely small sequences. Consequently,

$$(x_n \pm y_n) - (a \pm b) = \alpha_n \pm \beta_n.$$

According to Theorem 2.2, the sequence $\{\alpha_n \pm \beta_n\}$ is infinitely small. Thus the sequence $\{(x_n \pm y_n) - (a \pm b)\}$ is also infinitely small and therefore the sequence $\{x_n \pm y_n\}$ converges and has the number $a \pm b$ as its limit. ■

**Theorem 2.8.** *The product of convergent sequences $\{x_n\}$ and $\{y_n\}$ is a convergent sequence whose limit is equal to the product of the limits of the sequences $\{x_n\}$ and $\{y_n\}$.*

*Proof.* Let $a$ and $b$ be the limits of the sequences $\{x_n\}$ and $\{y_n\}$ respectively. Then, from formula (2), we find that

$$x_n = a + \alpha_n, \quad y_n = b + \beta_n,$$

where $\{\alpha_n\}$ and $\{\beta_n\}$ are infinitely small sequences. Consequently,

$$x_n y_n - ab = a\beta_n + b\alpha_n + \alpha_n\beta_n.$$

According to Theorems 2.2-2.4, the sequence $\{a\beta_n + b\alpha_n + \alpha_n\beta_n\}$ is infinitely small. Thus the sequence $\{x_n y_n - ab\}$ is also infinitely small and therefore the sequence $\{x_n y_n\}$ converges and has the number $ab$ as its limit. ■

**Theorem 2.9.** *The quotient of two convergent sequences $\{x_n\}$ and $\{y_n\}$, under the condition that the limit of $\{y_n\}$ is nonzero*, is a con-*

* By virtue of the condition $\lim_{n\to\infty} y_n \neq 0$, beginning with a certain number $N$, the elements $y_n$ do not vanish and therefore the quotient $\{x_n/y_n\}$ has sense for all $n > N$.

*vergent sequence whose limit is equal to the quotient of the limits of the sequences* $\{x_n\}$ *and* $\{y_n\}$.

*Proof.* Let $a$ and $b$ ($b \neq 0$) be the limits of the sequences $\{x_n\}$ and $\{y_n\}$ respectively. Then, from formula (2), we have

$$x_n = a + \alpha_n, \quad y_n = b + \beta_n,$$

where $\{\alpha_n\}$ and $\{\beta_n\}$ are infinitely small sequences. Consequently,

$$\frac{x_n}{y_n} - \frac{a}{b} = \frac{bx_n - ay_n}{by_n} = \frac{b(a+\alpha_n) - a(b+\beta_n)}{by_n} = \frac{1}{y_n}\left(\alpha_n - \frac{a}{b}\beta_n\right).$$

By virtue of the properties of infinitely small sequences, the sequence $\left\{\alpha_n - \frac{a}{b}\beta_n\right\}$ is infinitely small. We shall show that $\{1/y_n\}$ is a bounded sequence. Since $y_n \to b$ as $n \to \infty$, there is a number $N$, for $\varepsilon = |b|/2$, such that $|y_n - b| < |b|/2$ for all $n > N$. Therefore

$$|y_n| = |b - (b - y_n)| \geqslant |b| - |y_n - b| > |b| - \frac{|b|}{2} = \frac{|b|}{2},$$

i.e. $|y_n| > |b|/2$ and, consequently, $|1/y_n| < 2/|b|$ for all $n > N$, and this means that the sequence $\{1/y_n\}$ is bounded. According to Theorem 2.4, the sequence $\left\{\frac{1}{y_n}\left(\alpha_n - \frac{a}{b}\beta_n\right)\right\}$ is infinitely small and therefore the sequence $\left\{\frac{x_n}{y_n} - \frac{a}{b}\right\}$ is also infinitely small. Consequently, the sequence $\{x_n/y_n\}$ converges and has the number $a/b$ as its limit. ■

The theorems we have proved in this section are of great significance, both theoretical and practical.

**Example.** Let us find $\lim\limits_{n\to\infty} \frac{2n^2+n+1}{3n^2-1}$.

As $n \to \infty$, the numerator and denominator of the fraction tend to infinity and, consequently, we cannot use the theorem on the limit of a quotient since the hypothesis of this theorem assumes the existence of finite limits. We therefore first transform the given sequence dividing the numerator and denominator by $n^2$. Then, applying the theorem on the limit of a quotient and on the limit of a sum, we obtain

$$\lim_{n\to\infty} \frac{2n^2+n+1}{3n^2-1} = \lim_{n\to\infty} \frac{2+1/n+1/n^2}{3-1/n^2}$$

$$= \frac{\lim\limits_{n\to\infty}(2+1/n+1/n^2)}{\lim\limits_{n\to\infty}(3-1/n^2)} = \frac{\lim\limits_{n\to\infty} 2 + \lim\limits_{n\to\infty}(1/n) + \lim\limits_{n\to\infty}(1/n^2)}{\lim\limits_{n\to\infty} 3 - \lim\limits_{n\to\infty}(1/n^2)} = \frac{2+0+0}{3-0} = \frac{2}{3}.$$

**3. The limiting process in inequalities.**

**Theorem 2.10.** *If, beginning with a certain number, the elements of the convergent sequence $\{x_n\}$ satisfy the inequality $x_r \geqslant b$ ($x_n \leqslant b$), then the limit $a$ of this sequence also satisfies the inequality $a \geqslant b$ ($a \leqslant b$).*

*Proof.* Assume that, beginning with a certain number, all elements $x_n$ satisfy the inequality $x_n \geqslant b$. We have to prove the inequality $a \geqslant b$. We assume the contrary, i.e. that $a < b$.

Since $a$ is the limit of $\{x_n\}$, it follows that for $\varepsilon = b - a$ there is a number $N$ such that for $n > N$, there holds an inequality $|x_n - a| < b - a$ which is equivalent to the following two inequalities: $-(b-a) < x_n - a < b - a$. From the right-hand inequality we find that $x_n < b$ for $n > N$ and this contradicts the hypothesis. Consequently, $a \geqslant b$.

The case $x_n \leqslant b$ can be treated by analogy. ■

**Corollary 1.** *If, beginning with a certain number, the elements of the convergent sequences $\{x_n\}$ and $\{y_n\}$ satisfy the inequality $x_n \leqslant y_n$, then their limits satisfy the inequality $\lim\limits_{n\to\infty} x_n \leqslant \lim\limits_{n\to\infty} y_n$.*

Indeed, beginning with a certain number, the elements of the sequence $\{y_n - x_n\}$ are nonnegative and therefore its limit, i.e. $\lim\limits_{n\to\infty} (y_n - x_n) = \lim\limits_{n\to\infty} y_n - \lim\limits_{n\to\infty} x_n \geqslant 0$, is also nonnegative. Hence it follows that $\lim\limits_{n\to\infty} x_n \leqslant \lim\limits_{n\to\infty} y_n$.

**Corollary 2.** *If all elements of the convergent sequence $\{x_n\}$ are on the interval $[a, b]$, then its limit $c$ is also on this interval.*

Indeed, since $a \leqslant x_n \leqslant b$, it follows that $a \leqslant c \leqslant b$.

The following theorem plays a significant part in various applications.

**Theorem 2.11.** *Consider three sequences $\{x_n\}$, $\{y_n\}$, $\{z_n\}$ with $x_n \leqslant y_n \leqslant z_n$ for all $n$, and let the sequences $\{x_n\}$ and $\{z_n\}$ have the same limit $a$. Then the sequence $\{y_n\}$ also has a limit $a$.*

*Proof.* We take an arbitrary $\varepsilon > 0$. From this $\varepsilon$ we find, for the sequence $\{x_n\}$, a number $N_1$ such that $|x_n - a| < \varepsilon$ for $n > N_1$, i.e.

$$\underline{a - \varepsilon < x_n} < a + \varepsilon. \tag{3}$$

From the same $\varepsilon$ we find, for the sequence $\{z_n\}$, a number $N_2$ such that $|z_n - a| < \varepsilon$ for $n > N_2$, i.e.

$$a - \varepsilon < \underline{z_n < a + \varepsilon}. \tag{4}$$

Let $N = \max\{N_1, N_2\}$. Then inequalities (3) and (4) are simultaneously satisfied for $n > N$. Using the underlined inequalities as well as the inequalities given in the hypothesis, we obtain

$$a - \varepsilon < x_n \leqslant y_n \leqslant z_n < a + \varepsilon \quad \text{for} \quad n > N.$$

Hence

$$a - \varepsilon < y_n < a + \varepsilon \quad \text{or} \quad |y_n - a| < \varepsilon \quad \text{for} \quad n > N.$$

This means that the limit of the sequence $\{y_n\}$ is $a$. ■

### 1.2.3. Monotone Sequences

**1. The definition and the test for the convergence of monotone sequences.**

**Definition.** *The sequence* $\{x_n\}$ *is said to be increasing if* $x_n < x_{n+1}$ *for all* $n$, *nondecreasing if* $x_n \leqslant x_{n+1}$ *for all* $n$, *decreasing if* $x_n > x_{n+1}$ *for all* $n$, *and nonincreasing if* $x_n \geqslant x_{n+1}$ *for all* $n$.

All sequences of these kinds are united by a common title of *monotone sequences.* Increasing and decreasing sequences are also called *strictly monotone sequences.*

Here are some examples of monotone sequences.

1. The sequence $1, 1/2, 1/3, \ldots, 1/n, \ldots$ is decreasing and bounded.

2. The sequence $1, 1, 1/2, 1/2, 1/3, 1/3, \ldots, 1/n, 1/n, \ldots$ is nonincreasing and bounded.

3. The sequence $1, 2, 3, \ldots, n, \ldots$ is increasing and unbounded.

4. The sequence $1, 1, 2, 2, 3, 3, \ldots, n, n, \ldots$ is nondecreasing and unbounded.

5. The sequence $1/2, 2/3, 3/4, \ldots, n/(n+1), \ldots$ is increasing and bounded.

Note that monotone sequences are bounded, they have at least one bound: nondecreasing sequences are bounded from below ($x_n \geqslant x_1$ for all $n$), nonincreasing sequences are bounded from above ($x_n \leqslant x_1$ for all $n$). It turns out that if a monotone sequence is bounded from above and from below, i.e. is bounded, then it converges. Nonmonotone sequences do not possess this property. For example, the nonmonotone sequence $\{(-1)^n\}$ is bounded but does not converge (see the remark to Theorem 2.6).

We can cite the following fundamental theorem on monotone sequences.

**Theorem 2.12.** *A monotone bounded sequence converges.*

*Proof.* Let us consider the case of a nondecreasing sequence.

Assume that $x_n \leqslant x_{n+1}$ for all $n$ and there is a number $M$ such that all elements $x_n$ are not larger than $M$, i.e. $x_n \leqslant M$. We consider a number set $X$ which consists of elements of this sequence. By the hypothesis, this set is bounded from above and is nonempty. Therefore, by virtue of Theorem 1.1, the set $X$ has the least upper bound. We shall designate it as $a$ and prove that $a$ is a limit of the sequence.

Since $a$ is the least upper bound of the set of elements of the sequence $\{x_n\}$, it follows, according to the property of the least upper bound, that for any $\varepsilon > 0$ there is a number $N$ such that $x_N > a - \varepsilon$. Since $\{x_n\}$ is a nondecreasing sequence, we have $x_n > a - \varepsilon$ for $n > N$. On the other hand, by the definition of the upper bound, $x_n \leqslant a < a + \varepsilon$ for all $n$. Thus for $n > N$ we get inequalities $a - \varepsilon < x_n < a + \varepsilon$, i.e. $|x_n - a| < \varepsilon$ for $n > N$. And this means that the number $a$ is a limit of the sequence $\{x_n\}$.

The case of a nonincreasing sequence can be treated by analogy. ■

**Remark.** The boundedness of a monotone sequence is the necessary and sufficient condition of convergence.

Indeed, if a monotone sequence is bounded, then it converges by virtue of Theorem 2.12, and if a monotone sequence converges, then it is bounded according to Theorem 2.6.

**2. The number e.** Let us consider a sequence $\{x_n\}$ with the general term $x_n = \left(1+\frac{1}{n}\right)^n$:

$$(1+1)^1,\ \left(1+\frac{1}{2}\right)^2,\ \ldots,\ \left(1+\frac{1}{n}\right)^n,\ \ldots .$$

We shall prove that it converges. For our purpose, it is sufficient to prove that the sequence $\{x_n\}$ is increasing and is bounded from above. Applying Newton's* binomial theorem (see formula (10) in 1.6.3, Item 4), we obtain

$$x_n = 1 + n\frac{1}{n} + \frac{n(n-1)}{2!}\frac{1}{n^2} + \frac{n(n-1)(n-2)}{3!}\frac{1}{n^3} + \ldots + \frac{n(n-1)(n-2)\ldots[n-(n-1)]}{n!}\frac{1}{n^n}.$$

We represent this expression in the form

$$x_n = 2 + \frac{1}{2!}\left(1-\frac{1}{n}\right) + \frac{1}{3!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right) + \ldots + \frac{1}{n!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{n-1}{n}\right). \tag{1}$$

Then we represent $x_{n+1}$ in the analogous form

$$x_{n+1} = 2 + \frac{1}{2!}\left(1-\frac{1}{n+1}\right) + \frac{1}{3!}\left(1-\frac{1}{n+1}\right)\left(1-\frac{2}{n+1}\right) + \ldots + \frac{1}{(n+1)!}\left(1-\frac{1}{n+1}\right)\left(1-\frac{2}{n+1}\right)\ldots\left(1-\frac{n}{n+1}\right).$$

Now we note that $\left(1-\frac{k}{n}\right) < \left(1-\frac{k}{n+1}\right)$ for $0 < k < n$. Therefore each term in the expression for $x_{n+1}$ is greater than the respective term in the expression for $x_n$ and, in addition, the expression for $x_{n+1}$ includes one more positive term as compared to the expression for $x_n$. Consequently, $x_n < x_{n+1}$, i.e. the sequence $\{x_n\}$ is increasing.

To prove that the sequence is bounded from above, we note that each expression in parentheses in relation (1) is smaller than unity.

---

* Newton, Sir Isaac (1642-1727), superb English physicist, mathematician and astronomer.

We also take into account that $\frac{1}{n!} < \frac{1}{2^{n-1}}$ for $n > 2$ and obtain

$$x_n < 2 + \frac{1}{2!} + \frac{1}{3!} + \ldots + \frac{1}{n!} < 1 + 1 + \frac{1}{2} + \frac{1}{2^2} + \ldots + \frac{1}{2^{n-1}}.$$

Using the formula for the sum of a geometric progression, we arrive at an inequality

$$x_n < 1 + \frac{1 - 1/2^n}{1 - 1/2} = 3 - \frac{1}{2^{n-1}} < 3.$$

We have thus proved that the sequence $\{(1 + 1/n)^n\}$ is increasing and bounded from above. According to Theorem 2.12, it has a limit. We denote the limit by the letter e. Thus, by definition,

$$\mathrm{e} = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n.$$

Note that the number e plays a significant part in many problems of mathematics. In particular, it is the base of natural logarithms. In this section we only give the definition of the number e. Later on we shall consider the way of calculating this number with any degree of accuracy.

Here we only note that since $x_n < 3$ and since it can be immediately seen from (1) that $2 < x_n$, it follows that the number e is in the range $2 \leqslant \mathrm{e} \leqslant 3$. We have proved that the number e is irrational.

We shall prove a theorem which we shall use repeatedly when proving other theorems.

### 1.2.4. Theorem on Nested Intervals

Assume that there is a sequence of intervals $[a_1, b_1]$, $[a_2, b_2]$, ..., $[a_n, b_n]$, ... such that each successive interval is contained in the preceding one: $[a_1, b_1] \supset [a_2, b_2] \supset \ldots \supset [a_n, b_n] \supset \ldots$, i.e.

$$a_n \leqslant a_{n+1} < b_{n+1} \leqslant b_n \quad \text{for all } n \tag{1}$$

and let $\lim\limits_{n \to \infty} (b_n - a_n) = 0$. We call this sequence the *sequence of nested intervals.*

**Theorem 2.13.** *For any sequence of nested intervals there is a unique point which belongs to all intervals of that sequence.*

*Proof.* It follows from inequality (1) that the left-hand endpoints of the intervals form a nondecreasing sequence

$$a_1 \leqslant a_2 \leqslant a_3 \leqslant \ldots \leqslant a_n \leqslant a_{n+1} \leqslant \ldots \tag{2}$$

and the right-hand endpoints form a nonincreasing sequence

$$b_1 \geqslant b_2 \geqslant b_3 \geqslant \ldots \geqslant b_n \geqslant b_{n+1} \geqslant \ldots . \tag{3}$$

Sequence (2) is bounded from above and sequence (3) from below since $a_n \leqslant b_1$ and $b_n \geqslant a_1$ for any $n$. Consequently, by virtue of Theorem 2.12, these sequences have limits. Let $\lim\limits_{n\to\infty} a_n = c'$ and $\lim\limits_{n\to\infty} b_n = c''$. Then, from the condition

$$\lim_{n\to\infty} (b_n - a_n) = \lim_{n\to\infty} b_n - \lim_{n\to\infty} a_n = c'' - c' = 0,$$

it follows that $c' = c''$, i.e. the sequences $\{a_n\}$ and $\{b_n\}$ have a common limit. Denoting this limit by the letter $c$, we find that the inequalities $a_n \leqslant c \leqslant b_n$ hold true for any $n$, i.e. the point $c$ belongs to all intervals of sequence (1).

We shall prove now that this point is unique. We assume that one more point, $c_1$ ($c_1 \neq c$), belongs to all intervals of sequence (1). Then the inequality $b_n - a_n \geqslant |c_1 - c|$ must be satisfied for any $n$ and, consequently, $\lim\limits_{n\to\infty} (b_n - a_n) \geqslant |c_1 - c| \neq 0$, but this contradicts the hypothesis. ■

**Remark.** The theorem will not be true if instead of closed intervals we consider open ones. For instance, for the sequence of nested open intervals

$$(0,\ 1) \supset (0,\ 1/2) \supset (0,\ 1/4) \supset \ldots \supset (0,\ 1/2^n) \supset \ldots \qquad (4)$$

there is not a single point which would belong to all intervals. Indeed, whichever point $c$ from the open interval (0, 1) we take, there is always a number $N$ such that $1/2^n < c$ for $n > N$ and, consequently, the point $c$ will not belong to the intervals of sequence (4) beginning with the interval $(0,\ 1/2^{N+1})$.

For what follows we shall need some knowledge from analytic geometry, and therefore the next chapter is devoted to this division of mathematics.

## Chapter 3

## PLANE ANALYTIC GEOMETRY

Analytic geometry is the division of mathematics which studies geometric images by algebraic methods. As early as the seventeenth century, the French mathematician Descartes* worked out the method of coordinates which is the apparatus of analytic geometry.

The method of coordinates is based on the concept of a coordinate system. We shall get acquainted with the rectangular (Cartesian) and polar systems of coordinates.

---

* Descartes, René (1596-1650), a philosopher and mathematician.

### 1.3.1. Rectangular System of Coordinates

Two mutually perpendicular axes, $Ox$ and $Oy$, which have a common origin $O$ and the same scale unit (Fig. 8) form a *rectangular system of coordinates on the plane.*

The $x$-axis is called the *abscissa axis* and the $y$-axis, the *axis of ordinates*, and the two axes together are called the *axes of coordinates.*

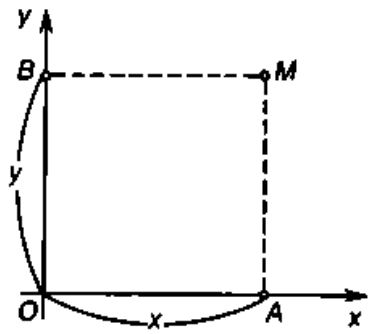



Fig. 8

The point $O$ of intersection of the axes is the *origin of coordinates.* The plane on which the $x$-axis and $y$-axis lie is known as the *coordinate plane* and is designated as $Oxy$.

Let $M$ be an arbitrary point of the plane. Let us draw from it perpendiculars $MA$ and $MB$ to the $x$-axis and $y$-axis.

The values $OA$ and $OB$ of the directed segments $\overrightarrow{OA}$ and $\overrightarrow{OB}$ are the *rectangular coordinates $x$ and $y$ of the point $M$*, i.e. $x = OA$ and $y = OB$.

The coordinates $x$ and $y$ of the point $M$ are its *abscissa* and *ordinate* respectively. The fact that the point $M$ has the coordinates $x$ and $y$ is symbolically designated as $M(x, y)$, the first letter in the parentheses denoting the abscissa and the second, the ordinate. The origin has coordinates (0, 0).

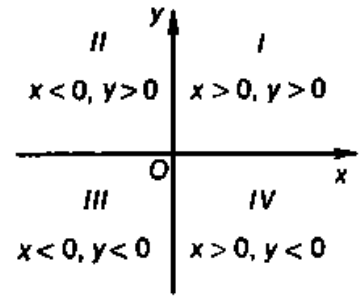



Fig. 9

Thus, with the chosen system of coordinates, every point $M$ of the plane is put into correspondence with a single pair $(x, y)$*, the numbers $x$ and $y$ being its rectangular coordinates, and, conversely, every pair of numbers $(x, y)$ is associated with a point $M$ on the $xy$-plane such that its abscissa is equal to $x$ and its ordinate to $y$, and this point is unique.

Thus the introduction of the rectangular system of coordinates on the plane makes it possible to establish a one-to-one correspondence between the set of all points of the plane and the set of pairs of numbers, and this permits using algebraic methods in solving geometry problems.

The coordinate axes divide the plane into four parts which are called *quadrants*, or *coordinate angles*, and are labelled by the Roman

* We speak here of an ordered pair of numbers, i.e. of the set of two numbers in which it is indicated which of the numbers is the first and which is the second. If $x \neq y$, then the pairs $(x, y)$ and $(y, x)$ are distinct since in the first of them the first number is $x$ and in the second, $y$.

numerals $I$, $II$, $III$ and $IV$ as shown in Fig. 9. This figure also shows the signs of coordinates of points depending on their location in a quadrant.

### 1.3.2. Simple Problems of Plane Analytic Geometry

**1. The distance between two points.**

**Theorem 3.1.** *For any two points $M_1$ $(x_1, y_1)$ and $M_2$ $(x_2, y_2)$ of the plane the distance $d$ between them is expressed by the formula*

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}. \tag{1}$$

*Proof.* We draw perpendiculars $M_1B$ and $M_2A$ from the points $M_1$ and $M_2$ to the $y$-axis and $x$-axis respectively and denote the

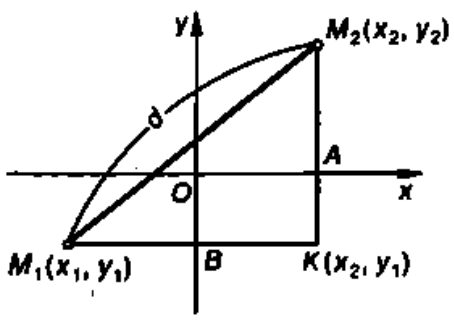



Fig. 10

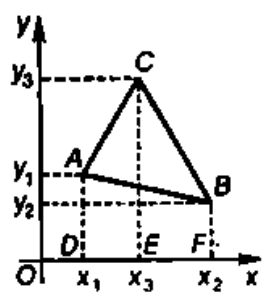



Fig. 11

point of the intersection of the straight lines $M_1B$ and $M_2A$ by $K$ (Fig. 10). The point $K$ has coordinates $(x_2, y_1)$ and therefore (see 1.1.3)

$$|M_1K| = |x_2 - x_1|, \quad |M_2K| = |y_2 - y_1|.$$

Since the triangle $M_1M_2K$ is right-angled, it follows, in accordance with the Pythagorean theorem, that

$$d = \sqrt{(M_1K)^2 + (M_2K)^2} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}. \quad \blacksquare$$

**2. The area of a triangle.**

**Theorem 3.2.** *For any points $A$ $(x_1, y_1)$, $B$ $(x_2, y_2)$, $C$ $(x_3, y_3)$, which do not lie on the same straight line, the area $s$ of the triangle $ABC$ is expressed by the formula*

$$s = \frac{1}{2} \left| [(x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1)] \right|. \tag{2}$$

*Proof.* The area of the triangle $ABC$ shown in Fig. 11 can be found as follows:

$$s_{ABC} = s_{ADEC} + s_{BCEF} - s_{ABFD}, \tag{3}$$

where $s_{ADEC}$, $s_{BCEF}$, $s_{ABFD}$ are the areas of the corresponding trapezoids. Since

$$s_{ADEC} = |DE| \frac{|AD| + |CE|}{2} = \frac{(x_3 - x_1)(y_3 + y_1)}{2},$$

$$s_{BCEF} = |EF| \frac{|EC| + |BF|}{2} = \frac{(x_2 - x_3)(y_2 + y_3)}{2},$$

$$s_{ABFD} = |DF| \frac{|AD| + |BF|}{2} = \frac{(x_2 - x_1)(y_1 + y_2)}{2},$$

we can substitute the expressions for these areas into relation (3) and get a formula

$$s = \frac{1}{2} \left| [(x_1 - x_2)(y_1 + y_2) + (x_2 - x_3)(y_2 + y_3) + (x_3 - x_1)(y_3 + y_1)] \right|,$$

which yields formula (2). By analogy we can prove formula (2) for any other position of the triangle $ABC$. ■

**Example.** Given points $A$ (1, 1), $B$ (6, 4), $C$ (8, 2), find the area of the triangle $ABC$. From formula (2) we have

$$s = \frac{1}{2} \left| [(6-1)(2-1) - (8-1)(4-1)] \right| = \frac{1}{2} \left| [-16] \right| = 8.$$

**3. Division of a line segment in a given ratio.** Assume that an arbitrary segment $M_1M_2$ is given on the plane and $M$ is any point of this segment different from the point $M_2$ (Fig. 12).

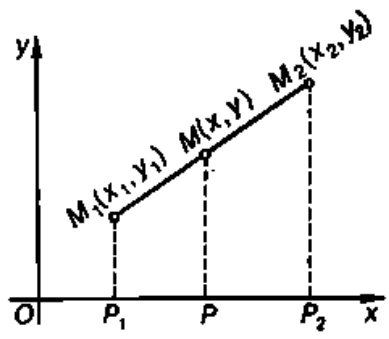


Fig. 12

The number $\lambda$ defined by the relation

$$\lambda = \frac{|M_1M|}{|MM_2|} \tag{4}$$

is the *ratio* in which the point $M$ divides the segment $M_1M_2$.

The problem of dividing a segment in a given ratio consists in finding the coordinates of the point $M$ from the given ratio $\lambda$ and the given coordinates of the points $M_1$ and $M_2$.

The following theorem makes it possible to solve this problem.

**Theorem 3.3.** *If the point $M$ $(x, y)$ divides the segment $M_1M_2$ in the ratio $\lambda$, then the coordinates of this point are defined by the formulas*

$$x = \frac{x_1 + \lambda x_2}{1 + \lambda}, \quad y = \frac{y_1 + \lambda y_2}{1 + \lambda}, \tag{5}$$

*where $(x_1, y_1)$ are the coordinates of the point $M_1$ and $(x_2, y_2)$ are the coordinates of the point $M_2$.*

*Proof.* Assume that the straight line $M_1M_2$ is not perpendicular to the $x$-axis. We drop perpendiculars from the points $M_1$, $M$, $M_2$ to the $x$-axis and denote the points of their intersection with the $x$-axis

by $P_1$, $P$, $P_2$ respectively (see Fig. 12). By virtue of the theorem on the proportionality of line segments included between parallel lines, considered in elementary geometry, we have

$$\frac{|P_1P|}{|PP_2|}=\frac{|M_1M|}{|MM_2|}=\lambda,$$

but $|P_1P| = |x - x_1|$ and $|PP_2| = |x_2 - x|$ (see 1.1.3). Since the numbers $(x - x_1)$ and $(x_2 - x)$ are of the same sign (they are positive for $x_1 < x_2$ and negative for $x_1 > x_2$), it follows that $\frac{|x-x_1|}{|x_2-x|} = \frac{x-x_1}{x_2-x}$. Therefore $\frac{x-x_1}{x_2-x} = \lambda$, whence $x = \frac{x_1+\lambda x_2}{1+\lambda}$. If the straight line $M_1M_2$ is perpendicular to the $x$-axis, then $x_1 = x_2 = x$ and this formula is evidently also valid. We have got the first of formulas (5).

We can obtain the second formula by analogy. ■

**Corollary.** If $M_1(x_1, y_1)$ and $M_2(x_2, y_2)$ are two arbitrary points and the point $M(x, y)$ is the midpoint of the segment $M_1M_2$, i.e. $|M_1M| = |MM_2|$, then $\lambda = 1$, and from formulas (5) we get

$$x=\frac{x_1+x_2}{2},\quad y=\frac{y_1+y_2}{2}.$$

Thus every coordinate of the midpoint of the segment is equal to half the sum of the respective coordinates.

**Example.** Given points $M_1(1, 1)$ and $M_2(7, 4)$, find the point $M(x, y)$ which is twice as close to $M_1$ as to $M_2$.

*Solution.* The required point $M$ divides the segment $M_1M_2$ in the ratio $\lambda = 1/2$. Using formulas (5), we find the coordinates of this point: $x = 3$ and $y = 2$.

### 1.3.3. Polar Coordinates

The *polar system of coordinates* is the second in importance after the rectangular system. It consists of a point $O$, called the *pole*, and a ray $OE$ emanating from it, which is the *polar axis*. In addition, a scale unit is specified to measure the lengths of segments.

Let a polar system of coordinates be given and assume that $M$ is an arbitrary point of the plane. Assume also that $\rho$ is the distance from the point $M$ to the point $O$ and $\varphi$ is the angle through which the polar axis must be turned to make it coincident with the ray $OM$ (Fig. 13).

The *polar coordinates of the point M* are the numbers $\rho$ and $\varphi$, the number $\rho$, being the first coordinate, called the *polar radius*, and the number $\varphi$, being the second coordinate, called the *polar angle*.

The point $M$ with the polar coordinates $\rho$ and $\varphi$ is designated as $M(\rho, \varphi)$. The polar radius can evidently have any nonnegative value, i.e. $0 \leqslant \rho < +\infty$. It is usually assumed that the polar angle varies in the range $0 \leqslant \varphi < 2\pi$. However, in a number of cases we

have to consider angles larger than $2\pi$ as well as negative angles, i.e. angles reckoned from the polar axis clockwise.

Let us establish a correspondence between the polar coordinates of a point and its rectangular coordinates. We shall assume that the origin of the rectangular system of coordinates is at the pole and the

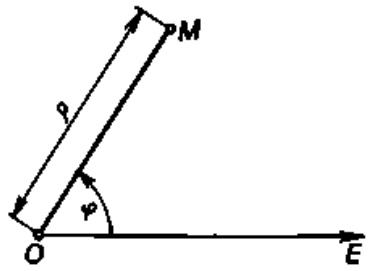



Fig. 13

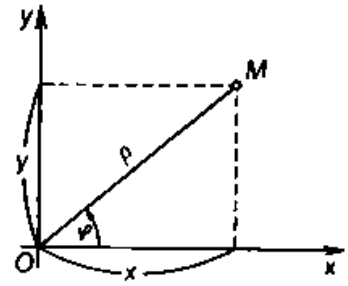



Fig. 14

positive semi-axis of abscissas coincides with the polar axis. Let the point $M$ have rectangular coordinates $x$ and $y$ and polar coordinates $\rho$ and $\varphi$ (Fig. 14). It is evident that

$$x = \rho \cos \varphi, \quad y = \rho \sin \varphi. \tag{1}$$

Formulas (1) express the rectangular coordinates in terms of the polar coordinates. The expressions for the polar coordinates in terms of the rectangular ones follow from formulas (1):

$$\rho = \sqrt{x^2 + y^2}, \quad \tan \varphi = \frac{y}{x}. \tag{2}$$

Note that the formula $\tan \varphi = y/x$ defines two values of the polar angle $\varphi$ since $\varphi$ varies from 0 to $2\pi$. From these two values of the angle $\varphi$ we choose that one for which formulas (1) are satisfied.

**Example.** Given rectangular coordinates (2, 2) of a point, find its polar coordinates assuming that the pole coincides with the origin of the rectangular system and the polar axis with the positive semi-axis of abscissas.

*Solution.* From formulas (2) we have

$$\rho = 2\sqrt{2}, \quad \tan \varphi = 1.$$

In accordance with the second of these relations, $\varphi = \pi/4$ or $\varphi = 5\pi/4$. But since $x = 2 > 0$ and $y = 2 > 0$, we must take $\varphi = \pi/4$.

### 1.3.4. Transformation of Rectangular Coordinates

When solving problems of analytic geometry, we must introduce, in addition to a given rectangular system of coordinates, some other rectangular systems. Then, naturally, the coordinates of points and the equations of curves change. A problem arises of how, knowing

the coordinates of a point in one system, to find the coordinates of that point in another system. The formulas for the transformation of coordinates make it possible to solve this problem. Let us consider two kinds of transformation of rectangular coordinates:

(1) *parallel displacement of axes*, when the position of the origin changes but the directions of the axes remain the same;

(2) *rotation of coordinate axes*, when both axes are rotated in the same direction and through the same angle and the origin remains fixed.

**1. Parallel displacement of the axes.** Let a point $M$ of the plane have coordinates $(x, y)$ in the rectangular system $Oxy$. We displace the origin to a point $O'(a, b)$, where $a$ and $b$ are the coordinates of the

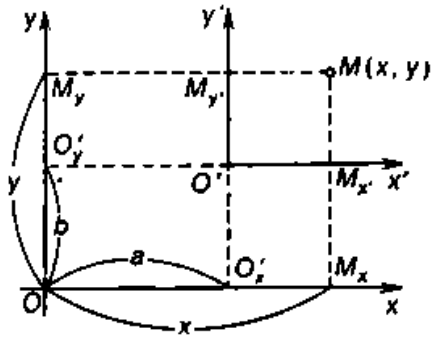



Fig. 15

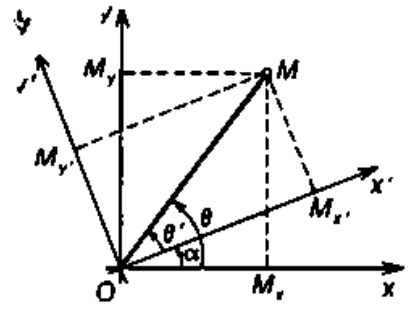



Fig. 16

new origin in the old system of coordinates $Oxy$. We make the new axes of coordinates $O'x'$ and $O'y'$ have the same direction as the old axes $Ox$ and $Oy$. We designate the coordinates of the point $M$ in the $O'x'y'$ system (the new coordinates) as $(x', y')$. Next we shall derive formulas which express a relationship between the new and old coordinates of the point $M$. For that purpose, we draw perpendiculars $MM_x \perp Ox$, $MM_y \perp Oy$, $O'O'_x \perp Ox$, $O'O'_y \perp Oy$ and introduce the designations $M_{x'}$ and $M_{y'}$ for the points of intersection of the straight lines $MM_x$ and $MM_y$ with the $O'x'$ and $O'y'$ axes respectively (Fig. 15). Then, using the fundamental identity (see 1.1.3), we obtain

$$x = OM_x = OO'_x + O'_xM_x = OO'_x + O'M_{x'} = a + x',$$
$$y = OM_y = OO'_y + O'_yM_y = OO'_y + O'M_{y'} = b + y'.$$

Thus

$$x = x' + a, \quad y = y' + b \tag{1}$$

or

$$x' = x - a, \quad y' = y - b. \tag{2}$$

These are precisely the formulas we have sought.

**2. Rotation of the coordinate axes.** Let us rotate the system of coordinates $Oxy$ about the origin $O$ through an angle $\alpha$ to the position $Ox'y'$ (Fig. 16). Assume that the point $M$ has coordinates $(x, y)$ in

the old system $Oxy$ and coordinates $(x', y')$ in the new system $Ox'y'$. We shall derive formulas which establish a relationship between the old and new coordinates of the point $M$. For that purpose, we designate the polar coordinates of the point $M$ as $(\rho, \theta)$ assuming the positive semi-axis $Ox$ to be the polar axis, and the polar coordinates of the same point $M$ as $(\rho, \theta')$ assuming the positive semi-axis $Ox'$ to be the polar axis. Evidently, in each case $\rho = |OM|$ and $\theta = \theta' + \alpha$. Furthermore, according to formulas (1) in 1.3.3,

$$x = \rho \cos \theta, \quad y = \rho \sin \theta$$

and, similarly,

$$x' = \rho \cos \theta', \quad y' = \rho \sin \theta'.$$

Thus

$$\begin{aligned} x = \rho \cos \theta &= \rho \cos (\theta' + \alpha) \\ &= \rho (\cos \theta' \cos \alpha - \sin \theta' \sin \alpha) \\ &= \rho \cos \theta' \cos \alpha - \rho \sin \theta' \sin \alpha \\ &= x' \cos \alpha - y' \sin \alpha, \end{aligned}$$

$$\begin{aligned} y = \rho \sin \theta &= \rho \sin (\theta' + \alpha) \\ &= \rho (\cos \theta' \sin \alpha + \sin \theta' \cos \alpha) \\ &= \rho \cos \theta' \sin \alpha + \rho \sin \theta' \cos \alpha \\ &= x' \sin \alpha + y' \cos \alpha. \end{aligned}$$

We thus have

$$\begin{cases} x = x' \cos \alpha - y' \sin \alpha, \\ y = x' \sin \alpha + y' \cos \alpha. \end{cases} \tag{3}$$

Expressing $x'$ and $y'$, which appear in these relations in terms of $x$ and $y$, we obtain

$$\begin{cases} x' = x \cos \alpha + y \sin \alpha, \\ y' = -x \sin \alpha + y \sin \alpha. \end{cases}$$

**Example.** Find the coordinates of the point $M$ (3, 5) in the new system of coordinates $O'x'y'$ whose origin $O'$ is at the point (−2, 1) and the axes are parallel to the axes of the old system of coordinates $Oxy$.

*Solution.* From formulas (2) we have

$$x' = 3 + 2 = 5, \quad y' = 5 - 1 = 4,$$

i.e. in the new system of coordinates the point $M$ has coordinates (5, 4).

### 1.3.5. The Equation of a Plane Curve

Let us consider an equation of the form

$$F(x, y) = 0, \tag{1}$$

which relates the variables $x$ and $y$. We call (1) an *equation of two variables* $x$ and $y$ if it is valid not for all pairs of numbers $x$ and $y$.

Examples of equations: $2x + 3y = 0$, $x^2 + y^2 - 25 = 0$, $\sin x + \sin y - 1 = 0$.

If equation (1) is valid for all pairs of numbers $x$ and $y$, then it is an *identity*.

Examples of identities: $(x + y)^2 - x^2 - 2xy - y^2 = 0$, $(x + y)(x - y) - x^2 + y^2 = 0$.

The concept of an *equation of a curve* is one of the most important concepts in analytic geometry. Let us consider a rectangular system of coordinates and a curve $L$ on the plane (Fig. 17).

**Definition.** *Equation* (1) *is the equation of the curve* $L$ *(in a given system of coordinates) if it is satisfied by the coordinates* $x$ *and* $y$ *of any point which lies on the curve* $L$ *and is not satisfied by the coordinates of any point which does not lie on this curve.*

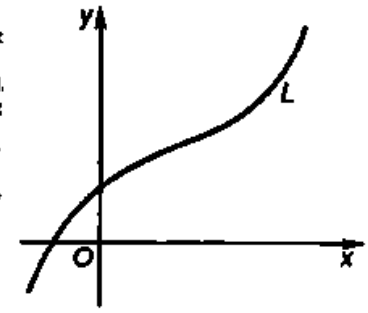


Fig. 17

It follows from the definition that the curve $L$ is the set of all points of the plane whose coordinates satisfy equation (1). We say that equation (1) defines (or specifies) the curve $L$.

The concept of an equation of a curve makes it possible to solve problems of geometry using algebraic methods. For instance, the problem of finding the point of intersection of two curves defined by the equations $x + y = 0$ and $x^2 + y^2 = 1$ reduces to an algebraic problem of solving the system of these equations.

The curve $L$ can be defined by an equation of the form

$$F(\rho, \varphi) = 0,$$

where $(\rho, \varphi)$ are the polar coordinates of the point.

Let us consider some examples of equations of curves.

(1) $x - y = 0$. Having written this equation in the form $y = x$, we infer that the set of points whose coordinates satisfy the given equation is the bisector of the first and third quadrants. And this is the curve defined by the equation $x - y = 0$ (Fig. 18).

(2) $x^2 - y^2 = 0$. Representing the equation in the form $(x - y)(x + y) = 0$, we infer that the set of points whose coordinates satisfy the given equation consists of two straight lines containing the bisectors of the four quadrants (Fig. 19).

(3) $x^2 + y^2 = 0$. The set of points whose coordinates satisfy this equation consists of one point (0, 0). In this case the equation defines, as we say, a singular curve.

(4) $x^2 + y^2 + 1 = 0$. Since the numbers $x^2$ and $y^2$ are nonnegative for any $x$ and $y$, it follows that $x^2 + y^2 + 1 > 0$. This means that

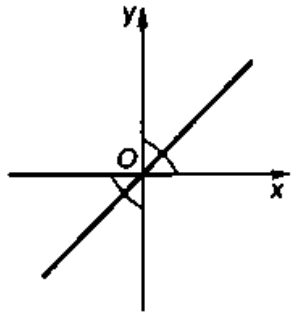



Fig. 18

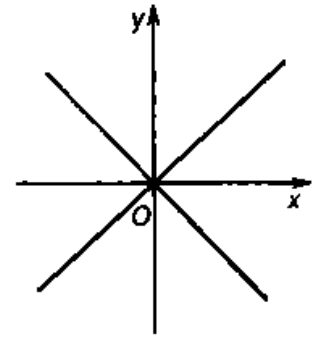



Fig. 19

there is not a single point whose coordinates satisfy the given equation, i.e. the equation does not define any geometric image on the plane.

(5) $\rho = a \cos \varphi$, where $a$ is a positive number and the variables $\rho$ and $\varphi$ are polar coordinates. We denote the point with polar coordi-

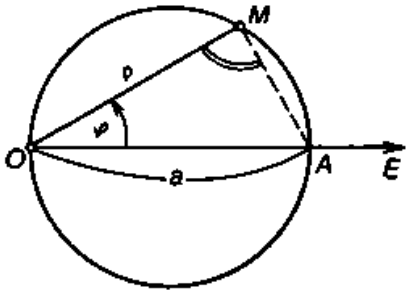



Fig. 20

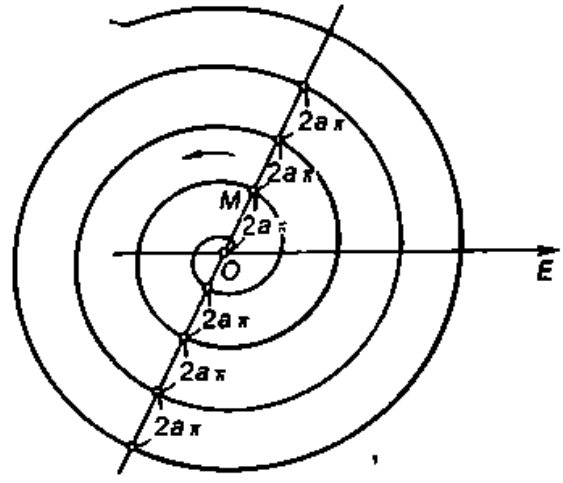



Fig. 21

nates $(\rho, \varphi)$ by $M$ and the point with polar coordinates $(a, 0)$ by $A$ (Fig. 20). If $\rho = a \cos \varphi$, where $0 < \varphi < \pi/2$, then the angle $OMA$ is a right angle, and vice versa. Consequently, the set of points whose polar coordinates satisfy the given equation is a circle with diameter $OA$.

(6) $\rho = a\varphi$, where $a$ is a positive number and $\rho$ and $\varphi$ are polar coordinates. We denote the point with polar coordinates $(\rho, \varphi)$ by $M$. If $\varphi = 0$, then $\rho = 0$ as well. If $\varphi$ increases, beginning with zero,

then $\rho$ increases in proportion to $\varphi$. Thus the point $M$ $(\rho, \varphi)$ emanates from the pole and moves about it with an increase in $\varphi$, simultaneously receding from it. The set of points whose polar coordinates satisfy the equation $\rho = a\varphi$ is known as the *spiral of Archimedes* (Fig. 21). In this case we suppose that $\varphi$ can assume any nonnegative values.

If the point $M$ makes one complete rotation about the pole, then $\varphi$ increases by $2\pi$ and $\rho$ by $2a\pi$, i.e. the spiral cuts any straight line which passes through the pole into equal segments (except for the segment which contains the pole) which are $2a\pi$ long.

In these examples we investigated the properties of a curve proceeding from the equation which defines it and thus found the type of the curve.

Let us now consider the converse problem: we have to find the equation of a set of points defined by some properties, i.e. of a given curve $L$.

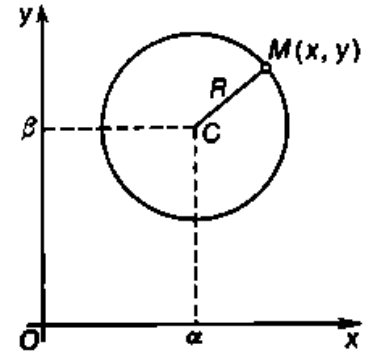


Fig. 22

**Example.** Derive the equation (in the given rectangular system of coordinates) of the set of points each of which is at the distance $R$ from the point $C$ $(\alpha, \beta)$. In other words, derive the equation of a circle of radius $R$ with centre at the point $C$ $(\alpha, \beta)$ (Fig. 22).

*Solution.* The distance from the arbitrary point $M$ $(x, y)$ to the point $C$ can be calculated by the formula $|MC| = \sqrt{(x-\alpha)^2 + (y-\beta)^2}$. If the point $M$ lies on the circumference of the circle, then $|MC| = R$, or $MC^2 = R^2$, i.e. the coordinates of the point $M$ satisfy the equation

$$(x-\alpha)^2 + (y-\beta)^2 = R^2. \tag{2}$$

But if the point $M$ $(x, y)$ does not lie on the circumference of the circle, then $MC^2 \neq R^2$, i.e. the coordinates of the point $M$ do not satisfy equation (2). Thus the required equation of the circle has form (2). Setting $\alpha = 0$ and $\beta = 0$ in (2), we get the equation of the circle of radius $R$ with centre at the origin, i.e. $x^2 + y^2 = R^2$.

### 1.3.6. First-Order Curves

**1. Slope-intercept form of the equation of a straight line.** Consider a straight line. We shall term as the *angle of inclination* of the straight line to the $x$-axis the angle $\alpha$ through which the $x$-axis must be rotated for its positive direction to coincide with one of the directions of the straight line. The angle $\alpha$ may have distinct values which differ from one another by $\pm n\pi$, where $n$ is a natural number. The least nonnegative value of the angle $\alpha$ through which the $x$-axis must be rotated counterclockwise for its positive direction to coincide

with one of the directions of the straight line is most often taken as the angle of inclination (Fig. 23). In this case $0 \leqslant \alpha < \pi$.

The tangent of the angle of inclination of a straight line to the $x$-axis is known as the *angular coefficient*, or *slope*, of the straight line and is denoted by the letter $k$:

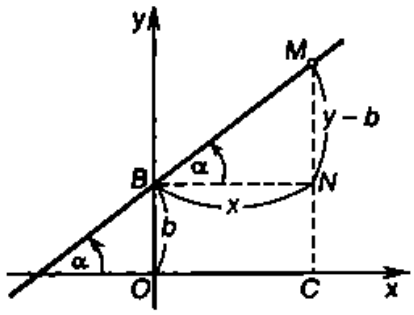


Fig. 23

$$k = \tan \alpha. \tag{1}$$

It follows from formula (1), in particular, that if $\alpha = 0$, i.e. the straight line is parallel to the $x$-axis, then $k = 0$. If $\alpha = \pi/2$, i.e. the straight line is perpendicular to the $x$-axis, then $k = \tan \alpha$ loses sense. In this case we say that the angular coefficient turns into infinity, or vanishes.

We shall derive the equation of the given straight line being given its angular coefficient $k$ and the value $b$ of the segment $OB$* which it intercepts on the $y$-axis (see Fig. 23) (i.e. the straight line is not perpendicular to the $x$-axis).

We denote by $M$ an arbitrary point of the plane with coordinates $x$ and $y$. If we draw straight lines $BN$ and $NM$ parallel to the axes, then in the case of $k \neq 0$ a right triangle $BNM$ is formed. Then the point $M$ lies on the straight line if and only if the quantities $NM$ and $BN$ satisfy the condition

$$\frac{NM}{BN} = \tan \alpha,$$

but $NM = CM - CN = CM - OB = y - b$ and $BN = x$. From this, taking formula (1) into account, we find that the point $M$ $(x, y)$ lies on the given straight line if and only if its coordinates satisfy the equation

$$\frac{y-b}{x} = k. \tag{2}$$

After transformation, equation (2) assumes the form

$$y = kx + b. \tag{3}$$

Equation (3) is known to be the *equation of a straight line in the slope-intercept form*. If $k = 0$, then the straight line is parallel to the $x$-axis and its equation has the form $y = b$.

Thus any curve which is not perpendicular to the $x$-axis has an equation of form (3). The converse is evidently also true, i.e. any

* Moreover, $b$ is the value of the directed segment $\overrightarrow{OB}$ on the $y$-axis. However, for the sake of brevity, we shall simply say "the value of the segment $OB$".

equation of form (3) defines a straight line which has an angular coefficient $k$ and intercepts a $b$-long segment on the $y$-axis.

**Example.** Construct a straight line defined by the equation $y = (3/4)\,x + 2$.

*Solution.* We lay off a segment $OB$ of length 2 on the $y$-axis (Fig. 24) and draw a segment whose value $BN = 4$ through the point $B$ parallel to the $x$-axis and a segment whose value $NM = 3$ through

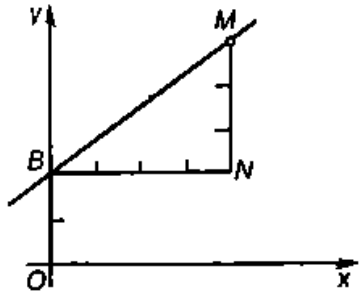



Fig. 24

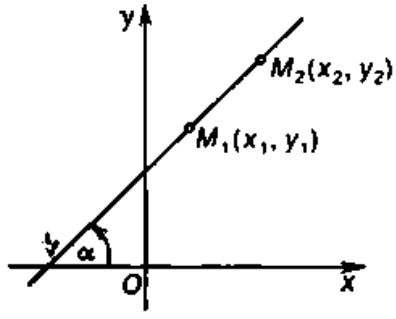



Fig. 25

the point $N$ parallel to the $y$-axis. Then we draw a straight line $BM$ which is the required line. It has an angular coefficient $k = 3/4$ and intercepts a segment $b = 2$ on the $y$-axis.

**2. Point-slope form of the equation of a straight line.** In some cases we have to set up an equation of a straight line being given one of its points $M_1\,(x_1, y_1)$ and its angular coefficient $k$. We write the equation of the straight line in form (3), where $b$ is an unknown number. Since the straight line passes through the point $M_1\,(x_1, y_1)$, the coordinates of this point satisfy equation (3), i.e. $y_1 = kx_1 + b$. Specifying $b$ in this relation and substituting it into equation (3), we get the required equation of the straight line:

$$y - y_1 = k\,(x - x_1). \tag{4}$$

**Remark.** If a straight line passes through the point $M_1\,(x_1, y_1)$ at right angles to the $x$-axis, i.e. its slope turns into infinity, then the equation of the straight line has the form $x - x_1 = 0$. Formally we can get it from (4) if we divide equation (4) by $k$ and then make $k$ tend to infinity.

**3. Two-point form of the equation of a straight line.** We are given two points $M_1\,(x_1, y_1)$ and $M_2\,(x_2, y_2)$ (Fig. 25). We write the equation of the straight line $M_1M_2$ in form (4), where $k$ is an unknown angular coefficient. Since the straight line $M_1M_2$ passes through the point $M_2$, the coordinates of this point satisfy equation (4): $y_2 - y_1 = k\,(x_2 - x_1)$. Specifying $k$ in this equation (provided that $x_1 \neq x_2$) and substituting it into equation (4), we get the required equation of the straight line:

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}\,(x - x_1).$$

Provided that $y_1 \neq y_2$, we can write this equation in the form

$$\frac{y-y_1}{y_2-y_1}=\frac{x-x_1}{x_2-x_1}. \tag{5}$$

If $y_1 = y_2$, then the equation of the required straight line has the form $y = y_1$. In this case the straight line is parallel to the $x$-axis. If $x_1 = x_2$, then the straight line which passes through the points $M_1$ and $M_2$ is parallel to the $y$-axis and its equation has the form $x = x_1$.

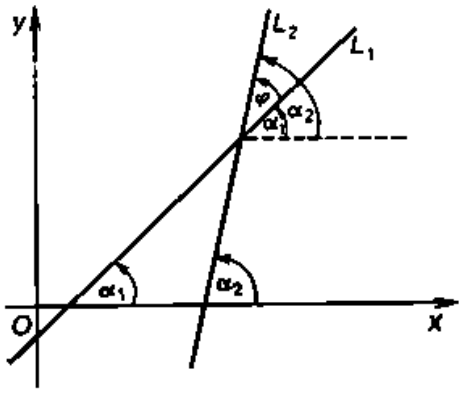



Fig. 26

**Example.** Set up an equation of a straight line which passes through the points $M_1$ (3, 1) and $M_2$ (5, 4).

*Solution.* Substituting the coordinates of the points $M_1$ and $M_2$ into relation (5), we get the required equation of the straight line:

$$\frac{x-3}{2}=\frac{y-1}{3}, \text{ or } 3x-2y-7=0.$$

**4. An angle between two straight lines.** Consider straight lines $L_1$ and $L_2$. Assume that the equation of $L_1$ has the form $y = k_1x + b_1$, where $k_1 = \tan \alpha_1$, and the equation of $L_2$ has the form $y = k_2x + b_2$, where $k_2 = \tan \alpha_2$ (Fig. 26). Let $\varphi$ be the angle between the lines $L_1$ and $L_2$: $0 \leqslant \varphi < \pi$.

From geometric considerations we establish the relationship between the angles $\alpha_1$, $\alpha_2$, $\varphi$: $\alpha_2 = \alpha_1 + \varphi$, or $\varphi = \alpha_2 - \alpha_1$. Hence

$$\tan \varphi = \tan(\alpha_2-\alpha_1)=\frac{\tan\alpha_2-\tan\alpha_1}{1+\tan\alpha_1\tan\alpha_2}$$

or

$$\tan \varphi = \frac{k_2-k_1}{1+k_2k_1} \tag{6}$$

Formula (6) defines one of the angles between the lines. The second angle is equal to $\pi - \varphi$.

**Example.** Two straight lines are defined by the equations $y = 2x + 3$ and $y = -3x + 2$. Find the angle between the lines.

*Solution.* Evidently $k_1 = 2$ and $k_2 = -3$ and therefore, from formula (6), we find that

$$\tan \varphi = \frac{-3-2}{1+(-3)\cdot 2}=\frac{-5}{-5}=1.$$

Thus one of the angles between the given lines is equal to $\pi/4$ and the other angle to $\pi - \pi/4 = 3\pi/4$.

**5. The conditions of parallelism and perpendicularity of two straight lines.** If the straight lines $L_1$ and $L_2$ are parallel, then $\varphi = 0$ and

$\tan\varphi = 0$. In this case the numerator on the right-hand side of formula (6) is zero, i.e. $k_2 - k_1 = 0$, whence

$$k_2 = k_1.$$

Thus *the condition of parallelism of two straight lines is the equality of their angular coefficients.*

If the straight lines $L_1$ and $L_2$ are perpendicular, i.e. $\varphi = \pi/2$, then $\alpha_2 = \pi/2 + \alpha_1$, $\tan\alpha_2 = \tan(\pi/2 + \alpha_1) = -\cot\alpha_1 = -1/(\tan\alpha_1)$, i.e.

$$k_2 = -\frac{1}{k_1}.$$

Thus *the condition of perpendicularity of two straight lines is that their angular coefficients are converse in value and opposite in sign.* This condition can be formally obtained from formula (6) by equating the denominator on the right-hand side of (6) to zero, which corresponds to the vanishing of $\tan\varphi$, i.e. to the equality $\varphi = \pi/2$.

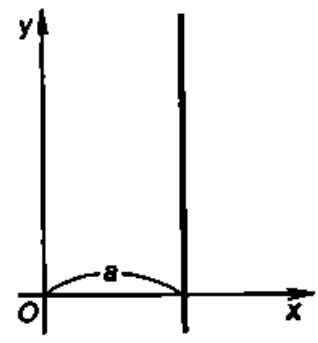


Fig. 27

**6. The general equation of a straight line.**

**Theorem 3.4.** *In a rectangular system of coordinates any straight line is defined by the linear equation*

$$Ax + By + C = 0 \qquad (7)$$

*and, conversely, for arbitrary coefficients $A$, $B$, $C$ ($A$ and $B$ are simultaneously nonzero) equation (7) defines a straight line in the rectangular system of coordinates $Oxy$.*

*Proof.* We shall first prove the first statement. If the straight line is not perpendicular to the $x$-axis, then, as was shown in Item 1, it has an equation $y = kx + b$, i.e. equation of form (7), where $A = k$, $B = -1$, $C = b$. If the straight line is perpendicular to the $x$-axis, then all its points have similar abscissas equal to the value $a$ of the segment intercepted by the straight line on the $x$-axis (Fig. 27). The equation of this straight line has the form $x = a$, i.e. is also a linear equation of form (7), where $A = 1$, $B = 0$, $C = -a$. We have thus proved the first statement.

Let us now prove the converse statement. Assume that we are given equation (7) in which at least one of the coefficients, $A$ or $B$, is nonzero.

If $B \neq 0$, then we can write (7) in the form

$$y = -\frac{A}{B}x - \frac{C}{B}.$$

Setting $k = -A/B$ and $b = -C/B$, we get an equation $y = kx + b$, i.e. an equation of form (3), which defines the straight line.

If $B = 0$, then $A \neq 0$ and equation (7) assumes the form $x = -C/A$. Denoting $-C/A$ by $a$, we get $x = a$, i.e. the equation of a straight line which is perpendicular to the $x$-axis. ■

Curves defined by the first-order equations in a rectangular system of coordinates are known as *first-order curves*. Thus every straight line is a first-order curve and, conversely, every first-order curve is a straight line.

An equation of the form $Ax + By + C = 0$ is the *general equation of a straight line*. It contains an equation of any straight line with the corresponding choice of the coefficients $A$, $B$, $C$.

**7. An incomplete first-order equation. The intercept equation of a straight line.** Let us consider three special cases when the equation $Ax + By + C = 0$ is *incomplete*, i.e. one of the coefficients is zero.

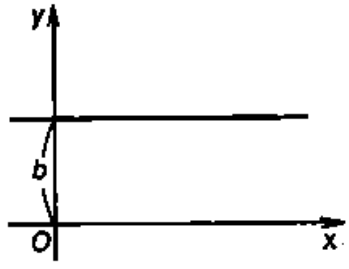



Fig. 28

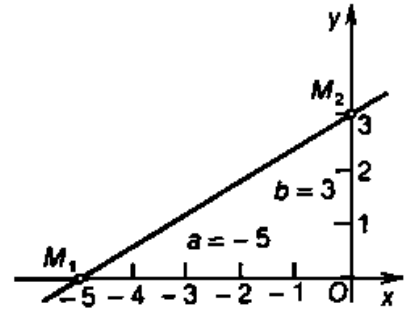



Fig. 29

(1) $C = 0$; the equation has the form $Ax + By = 0$ and defines the straight line which passes through the origin.

(2) $B = 0$ ($A \neq 0$); the equation has the form $Ax + C = 0$ and defines the straight line which is parallel to the $y$-axis. As was shown in Theorem 3.4, this equation can be reduced to the form $x = a$, where $a = -C/A$ and $a$ is the value of the segment which the straight line intercepts on the $x$-axis (see Fig. 27). In particular, if $a = 0$, then the straight line coincides with the $y$-axis. Thus the equation $x = 0$ defines the axis of ordinates.

(3) $A = 0$ ($B \neq 0$); the equation has the form $By + C = 0$ and defines the straight line which is parallel to the $x$-axis. This fact can be established by analogy with the preceding case. If we set $-C/B = b$, then the equation assumes the form $y = b$, where $b$ is the value of the segment which the straight line intercepts on the $y$-axis (Fig. 28). In particular, if $b = 0$, then the straight line coincides with the $x$-axis. Thus the equation $y = 0$ defines the abscissa axis.

Let us now be given an equation $Ax + By + C = 0$ under the condition that one of the coefficients, $A$, $B$, or $C$, is nonzero. We reduce it to the form

$$\frac{x}{-C/A} + \frac{y}{-C/B} = 1.$$

Introducing the designations $a = -C/A$ and $b = -C/B$, we obtain

$$\frac{x}{a} + \frac{y}{b} = 1. \tag{8}$$

Equation (8) is known as the *intercept equation of a straight line.* The numbers $a$ and $b$ are the values of the segments which the straight line intercepts on the coordinate axes. This form of the equation of a straight line is convenient for geometric construction of a straight line.

**Example.** A straight line is defined by the equation $3x - 5y + 15 = 0$. Derive an intercept equation for this straight line and construct the straight line.

*Solution.* For the given straight line the intercept equation has the form

$$\frac{x}{-5} + \frac{y}{3} = 1.$$

To construct this straight line, we lay off, on the coordinate axes $Ox$ and $Oy$, segments whose values are $a = -5$ and $b = 3$, respectively, and draw a straight line through the points $M_1$ $(-5, 0)$ and $M_2$ $(0, 3)$ (Fig. 29).

**8. Normal form of the equation of a straight line. The distance from a point to a straight line.** Consider a straight line $L$. We draw

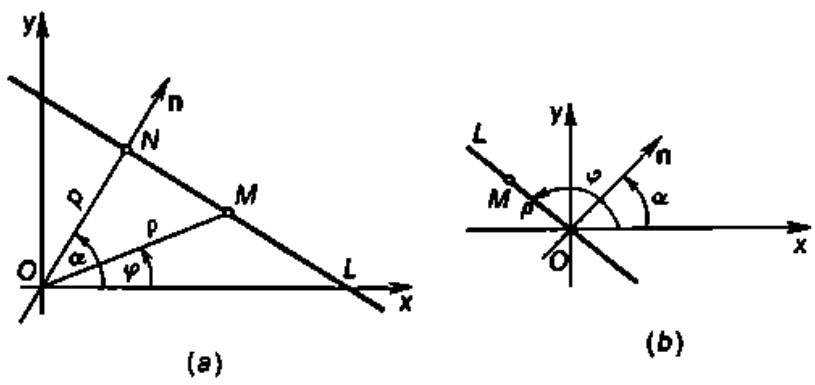



Fig. 30

a straight line **n** through the origin perpendicular to the given line and term it a *normal* to the straight line $L$. By the letter $N$ we mark the point at which the normal cuts the line $L$ (Fig. 30*a*). On the normal we mark the direction from the point $O$ to the point $N$. Thus the normal becomes an axis. If the points $N$ and $O$ coincide, then we take any one of the two possible directions as the direction of the normal.

We denote by $\alpha$ the angle through which we must rotate the $x$-axis counterclockwise to make its positive direction coincident with the direction of the normal, and by $p$, the length of the segment $ON$.

Thus $0 \leqslant \alpha < 2\pi$, $p \geqslant 0$. Next we derive the equation of the given straight line assuming the numbers $\alpha$ and $p$ to be known. For that purpose, we take, on the straight line, an arbitrary point $M$ with the polar coordinates $(\rho, \varphi)$, where $O$ is the pole and $Ox$ is the polar axis. If the points $O$ and $N$ do not coincide, then we have from the right triangle $ONM$

$$p = \rho \cos(\alpha - \varphi) = \rho(\cos\alpha \cos\varphi + \sin\alpha \sin\varphi).$$

We can rewrite this equation in the form

$$\rho \cos\varphi \cos\alpha + \rho \sin\varphi \sin\alpha - p = 0. \tag{9}$$

Since the points which do not lie on the straight line $L$ do not satisfy equation (9), it follows that (9) is the equation of the straight line $L$ in polar coordinates. From the formulas which relate the rectangular coordinates and the polar ones we have $\rho \cos\varphi = x$ and $\rho \sin\varphi = y$. Consequently, in the rectangular system of coordinates equation (9) assumes the form

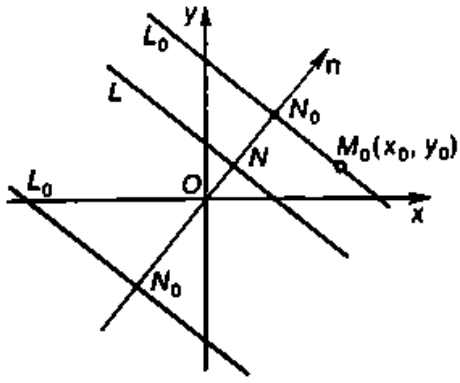


Fig. 31

$$x \cos\alpha + y \sin\alpha - p = 0. \tag{10}$$

If the points $O$ and $N$ coincide, then the straight line $L$ passes through the origin (Fig. 30*b*) and $p = 0$. In this case the equality $\cos(\varphi - \alpha) = 0$ is evidently satisfied for any point $M$ of the straight line $L$. Multiplying it by $\rho$, we find that $\rho \cos(\varphi - \alpha) = 0$, whence $\rho \cos\varphi \cos\alpha + \rho \sin\varphi \sin\alpha = 0$, or

$$x \cos\alpha + y \sin\alpha = 0.$$

Thus in this case as well we can represent the equation of the straight line in form (10).

Equation (10) is said to be the *equation of a straight line written in the normal form.*

Using this equation, we can determine the distance from a given point of the plane to a straight line.

Assume that $L$ is a straight line defined by the following equation written in the normal form: $x \cos\alpha + y \sin\alpha - p = 0$, and let $M_0(x_0, y_0)$ be a point not lying on this straight line. We have to find the distance $d$ from the point $M_0$ to the straight line $L$.

We draw a straight line $L_0$ through the point $M_0$ parallel to the straight line $L$. Let $N_0$ be the point of intersection of $L_0$ with the normal, and $p_0$, the length of the segment $ON_0$ (Fig. 31).

If the points $N$ and $N_0$ are on the same side of the point $O$, then the equation of the straight line $L_0$ written in the normal form is

$x \cos \alpha + y \sin \alpha - p_0 = 0$. Since the point $M_0(x_0, y_0) \in L_0$, it follows that $x_0 \cos \alpha + y_0 \sin \alpha - p_0 = 0$, whence $p_0 = x_0 \cos \alpha + y_0 \sin \alpha$. In this case

$$d = |p_0 - p| = |x_0 \cos \alpha + y_0 \sin \alpha - p|.$$

And if the points $N$ and $N_0$ are on different sides of the point $O$, then the equation of the straight line $L_0$ written in the normal form is $x \cos \alpha_1 + y \sin \alpha_1 - p_0 = 0$, where $\alpha_1$ differs from $\alpha$ by $\pi$. Consequently, $p_0 = x_0 \cos \alpha_1 + y_0 \sin \alpha_1 = -x_0 \cos \alpha - y_0 \sin \alpha$. In this case

$$d = |p_0 + p| = |-x_0 \cos \alpha - y_0 \sin \alpha + p|$$
$$= |x_0 \cos \alpha + y_0 \sin \alpha - p|.$$

Thus in each of the cases we have considered we get a formula

$$d = |x_0 \cos \alpha + y_0 \sin \alpha - p|. \tag{11}$$

Note that formula (11) is also suitable when the point $M_0(x_0, y_0)$ lies on the straight line $L$, i.e. its coordinates satisfy the equation of the straight line $L$: $x_0 \cos \alpha + y_0 \sin \alpha - p = 0$. In this case from formula (11) we get $d = 0$. It follows from formula (11) that to calculate the distance $d$ from the point $M_0$ to the straight line $L$, we must substitute the coordinates of the point $M_0$ for $(x, y)$ on the left-hand side of the equation of the straight line $L$ written in the normal form and take the absolute value of the number obtained.

We shall show now how to reduce the general equation of a straight line to the normal form. Let

$$Ax + By + C = 0 \tag{12}$$

be the general equation of a certain straight line and

$$x \cos \alpha + y \sin \alpha - p = 0 \tag{13}$$

be its normal form.

Since equations (12) and (13) define the same straight line, their coefficients are proportional. Multiplying all terms of equation (12) by an arbitrary factor $\mu \neq 0$, we get an equation

$$\mu A x + \mu B y + \mu C = 0.$$

With the requisite choice of $\mu$, the equation obtained turns into equation (13), i.e. the equalities

$$\mu A = \cos \alpha, \quad \mu B = \sin \alpha, \quad \mu C = -p \tag{14}$$

are satisfied. To find the factor $\mu$, we square the first two equalities and add them together. Then we get

$$\mu^2 (A^2 + B^2) = \cos^2 \alpha + \sin^2 \alpha = 1.$$

From this we obtain

$$\mu = \pm \frac{1}{\sqrt{A^2+B^2}}. \tag{15}$$

The number $\mu$ is called a *normalizing factor*. The sign of the normalizing factor is determined by means of the third relation (14). According to this relation, $\mu C$ is a negative number if $C \neq 0$. Consequently, in formula (15) we take the sign opposite to that of $C$. If $C = 0$, then the sign of the normalizing factor can be arbitrarily chosen.

Thus to reduce the general equation of a straight line to the normal form, we must find the value of the normalizing factor $\mu$ and then multiply all terms of the equation by $\mu$.

**Example.** Given a straight line $3x - 4y + 10 = 0$ and a point $M$ (4, 3), find the distance $d$ from the point $M$ to the straight line.

*Solution.* We reduce the equation to the normal form. For that purpose, we find the normalizing factor using formula (15):

$$\mu = \frac{-1}{\sqrt{3^2+4^2}} = -\frac{1}{5}.$$

Multiplying the given equation by $\mu$, we get an equation

$$-\frac{3}{5}x + \frac{4}{5}y - 2 = 0$$

in the normal form. Using formula (11), we find the required distance:

$$d = \left| -\frac{3}{5}\cdot 4 + \frac{4}{5}\cdot 3 - 2 \right| = |-2| = 2.$$

### 1.3.7. Second-Order Curves

We consider here three kinds of a broken line: the *ellipse*, the *hyperbola* and the *parabola* whose equations in the rectangular system of coordinates are second-order equations. Curves of that kind are called *second-order curves.*

**1. The ellipse.**

**Definition.** *The ellipse is the set of all points of the plane for which the sum of the distances from two given points, called foci, is a constant quantity, which is larger than the distance between the foci.*

We designate the foci of the ellipse as $F_1$ and $F_2$, the distance $|F_1F_2|$ between the foci as $2c$, and the sum of the distances from an arbitrary point of the ellipse to the foci as $2a$. By definition, $2a > 2c$, or $a > c$.

To derive the equation of the ellipse, we introduce a rectangular system of coordinates on the plane so that the foci of the ellipse lie on the abscissa axis and the origin divides the segment $F_1F_2$ in two. Then the foci of the ellipse have the coordinates $F_1$ $(-c, 0)$

and $F_2$ $(c, 0)$ (Fig. 32). We derive the equation of the ellipse in the chosen system of coordinates.

Let $M$ $(x, y)$ be an arbitrary point of the plane. We designate the distances from the point $M$ to the foci as $r_1$ and $r_2$ ($r_1 = |F_1M|$ and $r_2 = |F_2M|$). The numbers $r_1$ and $r_2$ are the *focal radii* of the point $M$. It follows from the definition of the ellipse that the point $M$ $(x, y)$ lies on the given ellipse if and only if

$$r_1 + r_2 = 2a. \qquad (1)$$

By formula (1) from 1.3.2, we find that

$$r_1 = \sqrt{(x+c)^2+y^2}, \quad r_2 = \sqrt{(x-c)^2+y^2}. \qquad (2)$$

Substituting these expressions into relation (1), we get

$$\sqrt{(x+c)^2+y^2} + \sqrt{(x-c)^2+y^2} = 2a. \qquad (3)$$

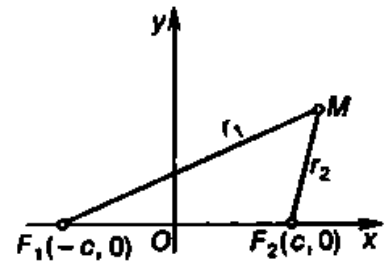


Fig. 32

Equation (3) is the required equation of the ellipse. However, it is inconvenient for practical applications and therefore the equation of the ellipse is usually reduced to a simpler form. We transfer the second radical to the right-hand side of the equation and then square both sides. We get

$$(x+c)^2+y^2 = 4a^2 - 4a\sqrt{(x-c)^2+y^2} + (x-c)^2+y^2,$$

or

$$a\sqrt{(x-c)^2+y^2} = a^2 - cx. \qquad (4)$$

We again square both sides of the equation

$$a^2x^2 - 2a^2cx + a^2c^2 + a^2y^2 = a^4 - 2a^2cx + c^2x^2$$

and obtain

$$(a^2 - c^2)\,x^2 + a^2y^2 = a^2\,(a^2 - c^2). \qquad (5)$$

We introduce into consideration a new quantity

$$b = \sqrt{a^2 - c^2} \qquad (6)$$

whose geometric meaning will be revealed later on. Since by the hypothesis $a > c$, it follows that $a^2 - c^2 > 0$ and, consequently, $b$ is a positive number. From relation (6) we have

$$b^2 = a^2 - c^2.$$

Therefore we can rewrite equation (5) in the form

$$b^2x^2 + a^2y^2 = a^2b^2.$$

Dividing both sides by $a^2b^2$, we finally obtain

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \qquad (7)$$

Since we obtained equation (7) from (3), the coordinates of any point of the ellipse, which satisfy equation (3), satisfy equation (7) as well. However, when we simplified equation (3), we twice squared its both sides and so extraneous roots may have appeared, as a result of which equation (7) may prove to be nonequivalent to equation (3). Let us ascertain that if the coordinates of a point satisfy equation (7), then they also satisfy equation (3), i.e. equations (3) and (7) are equivalent. To do that, it is evidently sufficient to show that for any point whose coordinates satisfy equation (7) the quantities $r_1$ and $r_2$ satisfy relation (1). Indeed, let the coordinates $x$ and $y$ of a certain point satisfy equation (7). Then, substituting the value $y^2 = b^2(1 - x^2/a^2)$ obtained from (7) into equation (2), we find, after simple transformations, that $r_1 = \sqrt{(a + cx/a)^2}$. Since $|x| \leqslant a$ (this follows from (7)) and $c/a < 1$, we have $a + cx/a > 0$ and therefore $r_1 = a + cx/a$.

By analogy we find that $r_2 = a - cx/a$. Adding these relations term by term, we get relation (1), and this is what we had to prove.

Thus, any point whose coordinates satisfy equation (7) belongs to the ellipse, and vice versa, i.e. equation (7) is the equation of the ellipse. Equation (7) is known as the *canonical* (or *simplest*) *equation of the ellipse*. Thus the ellipse is a second-order curve.

Let us now investigate the form of the ellipse using its canonical equation (7). Since equation (7) contains only terms with even degrees of the coordinates $x$ and $y$, the ellipse is symmetric with respect to the $x$-axis and $y$-axis and to the origin. We can thus find the form of the whole ellipse if we find the form of its part which lies in the first quadrant. For this part $y \geqslant 0$ and therefore, solving equation (7) for $y$, we get

$$y = \frac{b}{a}\sqrt{a^2 - x^2}. \tag{8}$$

Relation (8) yields the following statements:

(1) If $x = 0$, then $y = b$. Consequently, the point $(0, b)$ lies on the ellipse. We denote it by $B$.

(2) $y$ decreases when $x$ increases from 0 to $a$.

(3) If $x = a$, then $y = 0$. Consequently, the point $(a, 0)$ lies on the ellipse. We denote it by $A$.

(4) If $x > a$, then we get imaginary values of $y$. Consequently, there are no points of the ellipse whose $x > a$.

Thus the part of the ellipse which lies in the first quadrant is the arc $BA$* (Fig. 33).

Creating symmetry about the coordinate axes, we obtain the whole ellipse.

* In Chapter 6 we shall introduce the concept of the direction of convexity of the graph of the function $y = f(x)$ and show that the arc $BA$ is convex upwards.

**Remark.** If $a = b$, then equation (7) assumes the form $x^2 + y^2 = a^2$. This is the equation of a circle of radius $a$. Thus a circle is a special case of the ellipse. Note that we can get an ellipse from a circle of radius $a$ if we contract it $a/b$ times along the $y$-axis. Upon such a contraction, the point $(x, y)$ passes into a point $(x, y_1)$, where $y_1 = y\,(b/a)$. Substituting $y = y_1\,(a/b)$ into the equation of the circle, we get an equation of the ellipse

$$\frac{x^2}{a^2} + \frac{(y_1)^2}{b^2} = 1.$$

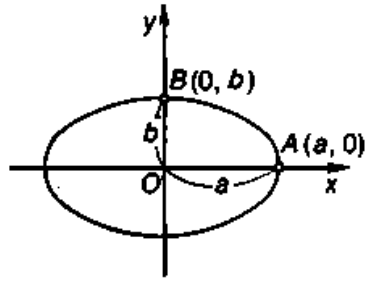


Fig. 33

The axes of symmetry are the *axes of the ellipse* and the centre of symmetry (the point of intersection of its axes) is the *centre of the ellipse*. The points at which the ellipse cuts the axes are its *vertices*. The vertices intercept segments equal to $2a$ and $2b$ on the axes. It follows from relation (6) that $a \geqslant b$. The quantities $a$ and $b$ are the *semimajor* and *semiminor axes of the ellipse* respectively. Accordingly, the axes of the ellipse are called the *major* and *minor axes*.

We shall introduce one more quantity which characterizes the shape of the ellipse.

**Definition.** *The eccentricity of the ellipse is the ratio $c/a$, where $c$ is half the distance between the foci and $a$ is the semimajor axis of the ellipse.*

The eccentricity is usually denoted by the letter $\varepsilon$: $\varepsilon = c/a$. Since $c < a$, it follows that $0 \leqslant \varepsilon < 1$, i.e. the eccentricity of the ellipse is smaller than unity. Taking into account that $c^2 = a^2 - b^2$, we find that

$$\varepsilon^2 = \frac{c^2}{a^2} = \frac{a^2 - b^2}{a^2} = 1 - \left(\frac{b}{a}\right)^2,$$

whence

$$\frac{b}{a} = \sqrt{1 - \varepsilon^2}.$$

From the last relation we can easily obtain the geometric interpretation of the eccentricity of the ellipse. For very small $\varepsilon$ the numbers $a$ and $b$ are almost equal, i.e. the ellipse is close to the circle. But if $\varepsilon$ is close to unity, then the number $b$ is rather small as compared to the number $a$, and the ellipse is very much elongated along the major axis. Thus the eccentricity of the ellipse characterizes the degree of elongation of the ellipse.

As is known, planets and some comets move along elliptical paths. The eccentricities of the orbits of planets prove to be very small and those of comets large, i.e. close to unity. Thus planets move almost along a circle and comets periodically approach the Sun

(the Sun is at one of the foci) and recede from it to a considerable distance.

**2. The hyperbola.**

**Definition.** *The hyperbola is the set of all points of the plane for which the absolute value of the difference of the distances from two given points, called foci, is a constant quantity, which is smaller than the distance between the foci.*

We designate the foci of the hyperbola as $F_1$ and $F_2$, the distance $|F_1F_2|$ between the foci as $2c$, and the absolute value of the difference of the distances from an arbitrary point of the hyperbola to the foci as $2a$. By definition, $2a < 2c$, or $a < c$.

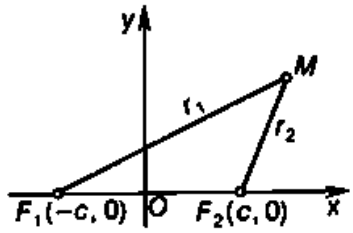



Fig. 34

To derive the equation of the hyperbola, we introduce a rectangular system of coordinates on the plane so that the foci of the hyperbola lie on the abscissa axis and the origin divides the segment $F_1F_2$ in two. Then the foci of the hyperbola have the coordinates $F_1(-c, 0)$ and $F_2(c, 0)$ (Fig. 34). We derive the equation of the hyperbola in the chosen system of coordinates.

Let $M(x, y)$ be an arbitrary point of the plane. The numbers $|F_1M|$ and $|F_2M|$ are the *focal radii* of the point $M$ and are designated as $r_1$ and $r_2$. It follows from the definition of the hyperbola that the point $M(x, y)$ lies on the given hyperbola if and only if $|r_1 - r_2| = 2a$. Hence

$$r_1 - r_2 = \pm 2a. \tag{9}$$

By formula (1) from 1.3.2, we find that

$$r_1 = \sqrt{(x+c)^2 + y^2}, \quad r_2 = \sqrt{(x-c)^2 + y^2}. \tag{10}$$

Substituting these expressions into relation (9), we obtain

$$\sqrt{(x+c)^2 + y^2} - \sqrt{(x-c)^2 + y^2} = \pm 2a. \tag{11}$$

Equation (11) is the required equation of the hyperbola. Let us simplify this equation as we did with equation (3) of the ellipse. We transfer the second radical to the right-hand side of the equation and the square both sides. We obtain

$$(x+c)^2 + y^2 = 4a^2 \pm 4a\sqrt{(x-c)^2 + y^2} + (x-c)^2 + y^2,$$

or

$$cx - a^2 = \pm a\sqrt{(x-c)^2 + y^2}. \tag{12}$$

We again square both sides of the equation

$$c^2x^2 - 2a^2cx + a^4 = a^2x^2 - 2a^2cx + a^2c^2 + a^2y^2$$

and get

$$(c^2 - a^2)\, x^2 - a^2y^2 = a^2\,(c^2 - a^2). \tag{13}$$

We introduce into consideration a new quantity

$$b = \sqrt{c^2 - a^2} \tag{14}$$

whose geometric meaning will be revealed later on. Since $c > a$, it follows that $c^2 - a^2 > 0$ and $b$ is a positive number. From relation (14) we have

$$b^2 = c^2 - a^2.$$

Equation (13) assumes the form

$$b^2x^2 - a^2y^2 = a^2b^2$$

or

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1. \tag{15}$$

As we did for the ellipse, we can prove that equations (15) and (11) are equivalent. Equation (15) is the *canonical equation of the hyperbola.*

We shall use the canonical equation of the hyperbola to investigate its form. Since equation (15) contains terms only with even degrees of the coordinates $x$ and $y$, the hyperbola is symmetric with respect to the $x$-axis and $y$-axis and to the origin. It is therefore sufficient to consider only the part of the hyperbola which lies in the first quadrant. For this part $y \geqslant 0$ and therefore, solving equation (15) for $y$, we get

$$y = \frac{b}{a}\sqrt{x^2 - a^2}. \tag{16}$$

Relation (16) yields the following statements:

(1) If $0 \leqslant x < a$, then $y$ assumes imaginary values, i.e. there are no points of the hyperbola with the abscissas $0 \leqslant x < a$.

(2) If $x = a$, then $y = 0$, i.e. the point $(a, 0)$ belongs to the hyperbola. We denote it by $A$.

(3) If $x > a$, then $y > 0$, with $y$ increasing with $x$ and $y \to +\infty$ as $x \to +\infty$.

Thus the variable point $M\,(x, y)$ on the hyperbola moves "to the right" and "upwards" with an increase in $x$, its initial position being $A\,(a, 0)$ (Fig. 35). Let us define more precisely how the point $M$ "recedes to infinity".

For that purpose, in addition to equation (16), we shall consider an equation

$$y = \frac{b}{a}\,x, \tag{17}$$

which defines a straight line with the slope $k = b/a$ passing through the origin. The part of this line, which lies in the first quadrant, is shown in Fig. 35. To construct it, we can use the right triangle $OAB$ with the legs $OA = a$ and $AB = b$.

We shall show that when the point $M$ recedes to infinity along the hyperbola, it infinitely approaches straight line (17), which is an *asymptote* to the hyperbola*.

We take an arbitrary value $x$ ($x \geqslant a$) and consider two points $M(x, y)$ and $N(x, Y)$, where

$$y = \frac{b}{a}\sqrt{x^2 - a^2}, \quad Y = \frac{b}{a}x.$$

The point $M$ lies on the hyperbola and the point $N$, on the straight line (17). Since the two points have the same abscissa $x$, the

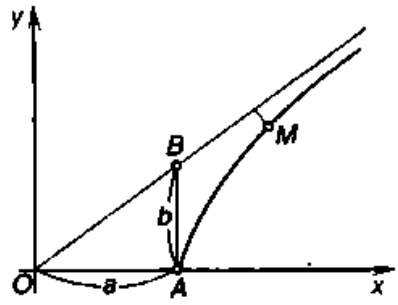



Fig. 35

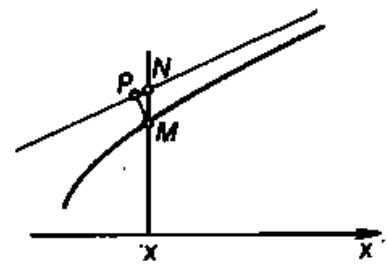



Fig. 36

straight line $MN$ is perpendicular to the $x$-axis (Fig. 36). We shall find the length of the segment $MN$.

Note first of all that for $x \geqslant a$

$$Y = \frac{b}{a}x = \frac{b}{a}\sqrt{x^2} > \frac{b}{a}\sqrt{x^2 - a^2} = y.$$

This means that for the same abscissa the point of the hyperbola lies under the respective point of the asymptote. Thus

$$|MN| = Y - y = \frac{b}{a}x - \frac{b}{a}\sqrt{x^2 - a^2} = \frac{b}{a}\left(x - \sqrt{x^2 - a^2}\right)$$

$$= \frac{b}{a}\,\frac{\left(x - \sqrt{x^2 - a^2}\right)\left(x + \sqrt{x^2 - a^2}\right)}{x + \sqrt{x^2 - a^2}} = \frac{ab}{x + \sqrt{x^2 - a^2}}.$$

It follows from the expression obtained that $|MN|$ tends to zero as $x \to +\infty$ since the denominator tends to $+\infty$ and the numerator is a constant quantity $ab$.

* In Chapter 6 we shall give the definition of the asymptote of the graph of the function $y = f(x)$ and show that the straight line $y = (b/a)x$ is an asymptote to the hyperbola. We shall also consider there the direction of convexity of the hyperbola.

We designate as $P$ the foot of the perpendicular dropped from the point $M$ to the straight line (17). Then $|MP|$ is the distance from the point $M$ to this line. Evidently, $|MP| < |MN|$, and since $|MN| \to 0$, it follows that $|MP| \to 0$ all the more so as $x \to +\infty$, i.e. the point $M$ infinitely approaches the straight line (17), and this is what we wished to prove.

We can easily establish now the shape of the whole hyperbola using the symmetry with respect to the coordinate axes (Fig. 37).

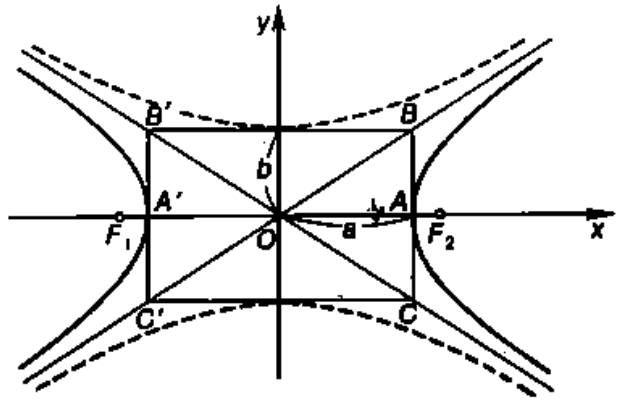



Fig. 37

The hyperbola consists of two branches (right and left) and has two asymptotes $y = (b/a)\,x$ and $y = -(b/a)\,x$, the first of which we have considered and the second is its symmetric reflection about the $x$-axis (or the $y$-axis).

The axes of symmetry are the *axes of the hyperbola* and the centre of symmetry (the point of intersection of the axes) is the *centre of the hyperbola*. One of the axes cuts the hyperbola at two points which are called its *vertices* (in Fig. 37 they are denoted by $A'$ and $A$). This is the *transverse (real) axis of the hyperbola*. The other axis has no points in common with the hyperbola and is known as the *conjugate axis of the hyperbola*. The rectangle $BB'C'C$ with sides $2a$ and $2b$ (see Fig. 37) is the *basic rectangle of the hyperbola*. The quantities $a$ and $b$ are the *semitransverse* and *semiconjugate axes of the hyperbola* respectively. The equation

$$\frac{y^2}{b^2} - \frac{x^2}{a^2} = 1$$

also defines the hyperbola. It is shown in Fig. 37 by dash lines and its vertices lie on the $y$-axis. This hyperbola is the *conjugate hyperbola* of (15). Both these hyperbolas have the same asymptotes.

The hyperbola with equal semi-axes ($a = b$) is *equilateral*, or *equiangular*, and its canonical equation has the form

$$x^2 - y^2 = a^2.$$

Since the basic rectangle of the equilateral hyperbola is a square,

the asymptotes to the equilateral hyperbola are perpendicular to each other.

**Definition.** *The eccentricity of the hyperbola is the ratio $c/a$, where $c$ is half the distance between the foci and $a$ is the semitransverse axis of the hyperbola.*

The eccentricity of the hyperbola (like that of the ellipse) is designated as $\varepsilon$. Since $c > a$, it follows that $\varepsilon > 1$, i.e. the eccentricity of the hyperbola is greater than unity. Noting that $c^2 = a^2 + b^2$, we find that

$$\varepsilon^2 = \frac{c^2}{a^2} = \frac{a^2+b^2}{a^2} = 1 + \left(\frac{b}{a}\right)^2,$$

whence

$$\frac{b}{a} = \sqrt{\varepsilon^2 - 1}.$$

From the last relation we can easily obtain the geometric interpretation of the eccentricity of the hyperbola.The smaller the eccentricity, i.e. the closer it is to unity, the smaller the ratio $b/a$, and this means that the basic rectangle is extended in the direction of the transverse axis. Thus the eccentricity of the hyperbola characterizes the shape of its basic rectangle and, hence, the shape of the hyperbola itself.

In the case of an equilateral hyperbola ($a = b$), $\varepsilon = \sqrt{2}$.

**3. The directrices of the ellipse and the hyperbola.**

**Definition 1.** *Two straight lines, which are perpendicular to the major axis of the ellipse, are symmetric with respect to the centre and lie at the distance $a/\varepsilon$ from it, are the directrices of the ellipse (here $a$ is the semimajor axis and $\varepsilon$ is the eccentricity of the ellipse).*

The equations of the directrices of the ellipse, which is defined by the canonical equation (7), have the form

$$x = -\frac{a}{\varepsilon}, \quad x = \frac{a}{\varepsilon}.$$

Since for the ellipse $\varepsilon < 1$, it follows that $a/\varepsilon > a$. Hence the right-hand directrix lies to the right of the right-hand vertex of the ellipses and the left-hand directrix lies to the left of the left-hand vertex (Fig. 38).

**Definition 2.** *Two straight lines, which are perpendicular to the transverse axis of the hyperbola, are symmetric with respect to the centre and lie at the distance $a/\varepsilon$ from it, are the directrices of the hyperbola (here $a$ is the semitransverse axis and $\varepsilon$ is the eccentricity of the hyperbola).*

The equations of the directrices of the hyperbola, which is defined by the canonical equation (15), have the form

$$x = -\frac{a}{\varepsilon}, \quad x = \frac{a}{\varepsilon}.$$

Since for the hyperbola $\varepsilon > 1$, it follows that $a/\varepsilon < a$. Hence the right-hand directrix lies between the centre and the right-hand vertex of the hyperbola and the left-hand directrix lies between the centre and the left-hand vertex (Fig. 39).

Using the concepts of the directrix and the eccentricity, we can formulate the general property inherent in the ellipse and the hyperbola. We have the following two theorems.

**Theorem 3.5.** *If $r$ is the distance from an arbitrary point $M$ of the ellipse to some focus and $d$ is the distance from the same point to the*

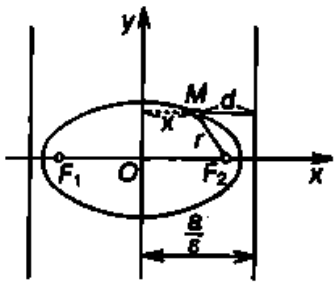



Fig. 38

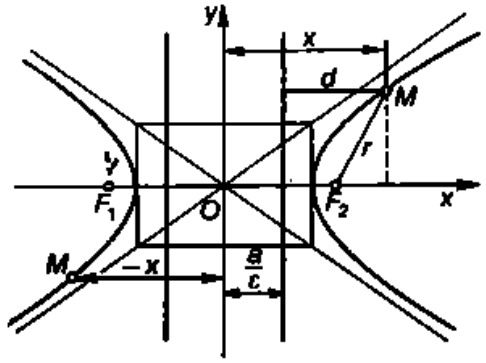



Fig. 39

*directrix, which corresponds to this focus, then the ratio $r/d$ is a constant quantity equal to the eccentricity of the ellipse.*

*Proof.* Let us assume, for the sake of definiteness, that we speak of the right-hand focus $F_2$ and the right-hand directrix. Let $M(x, y)$ be an arbitrary point of the ellipse (see Fig. 38). The distance from the point $M$ to the right-hand directrix is expressed by the relation

$$d = \frac{a}{\varepsilon} - x, \tag{18}$$

which can be easily derived using the figure. From relations (2) and (4) we have

$$r = r_2 = \sqrt{(x-c)^2 + y^2} = a - \frac{c}{a}x.$$

Setting $c/a = \varepsilon$, we get the formula for the distance from the point $M$ to the right-hand focus:

$$r = a - \varepsilon x. \tag{19}$$

From relations (18) and (19) we have

$$\frac{r}{d} = \frac{a - \varepsilon x}{\frac{a}{\varepsilon} - x} = \frac{(a - \varepsilon x)\varepsilon}{a - \varepsilon x} = \varepsilon. \ \blacksquare$$

**Theorem 3.6.** *If $r$ is the distance from an arbitrary point $M$ of the hyperbola to some focus and $d$ is the distance from the same point to the*

*directrix, which corresponds to this focus, then the ratio $r/d$ is a constant quantity equal to the eccentricity of the hyperbola.*

*Proof.* Let us assume, for the sake of definiteness, that we speak of the right-hand focus $F_2$ and the right-hand directrix. Let $M(x, y)$ be an arbitrary point of the hyperbola (see Fig. 39). We shall consider two cases.

(1) The point $M$ is on the right-hand branch of the hyperbola. Then the distance from the point $M$ to the right-hand directrix is expressed by the relation

$$d = x - \frac{a}{\varepsilon}, \tag{20}$$

which can be easily derived using the figure. From relations (10) and (12) we have

$$r = r_2 = \sqrt{(x-c)^2 + y^2} = \frac{c}{a}x - a.$$

Setting $c/a = \varepsilon$, we get the formula for the distance from the point $M$ to the right-hand focus:

$$r = \varepsilon x - a. \tag{21}$$

From relations (20) and (21) we have

$$\frac{r}{d} = \frac{\varepsilon x - a}{x - \frac{a}{\varepsilon}} = \frac{(\varepsilon x - a)\varepsilon}{\varepsilon x - a} = \varepsilon.$$

(2) The point $M$ is on the left-hand branch of the hyperbola. Then the distance from the point $M$ to the right-hand directrix is expressed by the relation (see Fig. 39)

$$d = -x + \frac{a}{\varepsilon}. \tag{22}$$

By analogy with (21), we can get the formula for the distance from the point $M$ to the right-hand focus:

$$r = -\left(\frac{c}{a}x - a\right) = -(\varepsilon x - a). \tag{23}$$

From relations (22) and (23) we have

$$\frac{r}{d} = \frac{-(\varepsilon x - a)}{-x + \frac{a}{\varepsilon}} = \frac{(-\varepsilon x + a)\varepsilon}{(-\varepsilon x + a)} = \varepsilon. \quad \blacksquare$$

The property of the ellipse and the hyperbola we have established can be used as the basis for the general definition of these curves: *the set of points for which the ratio of the distances to a focus and to the corresponding directrix is a constant quantity, equal to $\varepsilon$, is the ellipse if $\varepsilon < 1$ and the hyperbola if $\varepsilon > 1$.*

A question naturally arises what is the set of points defined in a similar way for $\varepsilon = 1$. It proves to be a new second-order curve known as a *parabola*.

**4. The parabola.**

**Definition.** *The parabola is the set of all points of the plane which are at the same distance from a given point, called a focus, and from a given straight line, called a directrix, and does not pass through the focus.*

To derive the equation of the parabola, we introduce a rectangular system of coordinates on the plane so that the abscissa axis passes through the focus at right angles to the directrix. We assume that the direction from the directrix to the focus is positive and put the origin at the midpoint between the focus and the directrix. Let us derive the equation of the parabola in the chosen system of coordinates.

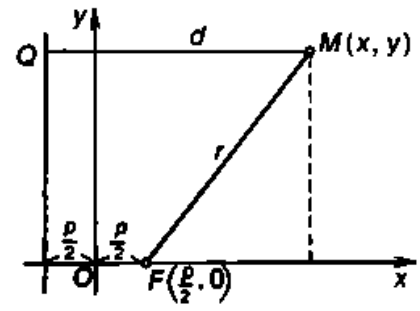


Fig. 40

We assume that $M(x, y)$ is an arbitrary point of the plane. We designate the distance from the point $M$ to the focus $F$ as $r$ ($r = |FM|$), the distance from the point $M$ to the directrix as $d$ and the distance from the focus to the directrix as $p$ (Fig. 40). The quantity $p$ is the parameter of the parabola whose geometric meaning will be revealed later on. The point $M$ lies on the given parabola if and only if

$$r = d. \tag{24}$$

The focus $F$ has the coordinates $(p/2, 0)$ and therefore, by formula (1) from 1.3.2, we find that

$$r = |FM| = \sqrt{\left(x - \frac{p}{2}\right)^2 + y^2}. \tag{25}$$

The distance $d$ is evidently expressed by the relation (see Fig. 40)

$$d = |MQ| = x + \frac{p}{2}. \tag{26}$$

Note that this formula only holds true for $x \geqslant 0$. But if $x < 0$, then, evidently, $r > d$ for the point $M(x, y)$ and, consequently, such a point does not lie on the parabola. Replacing $r$ and $d$ in relation (24) by their expressions (25) and (26), we find that

$$\sqrt{\left(x - \frac{p}{2}\right)^2 + y^2} = x + \frac{p}{2}. \tag{27}$$

And this is the required equation of the parabola. We shall reduce it to a more convenient form, for which purpose we square both

sides of relation (27). We obtain

$$x^2 - px + \frac{p^2}{4} + y^2 = x^2 + px + \frac{p^2}{4}$$

or

$$y^2 = 2px. \tag{28}$$

Let us verify whether equation (28), which results from the squaring of both sides of equation (27), acquired extraneous roots. For that purpose, it is sufficient to show that for any point $M(x, y)$, whose coordinates satisfy equation (28), relation (24) is satisfied. Indeed, it follows from equation (28) that $x \geqslant 0$ and therefore $d = p/2+x$ for the point $M(x, y)$ with a nonnegative abscissa. Substituting the value of $y^2$ from (28) into expression (25) for $r$ and taking into account that $x \geqslant 0$, we get $r = p/2+x$, i.e. the quantities $r$ and $d$ are equal, and that is what we wished to prove. Thus equation (28) is satisfied by the coordinates of the points of the given parabola and only by those points, i.e. equation (28) is the equation of the given parabola.

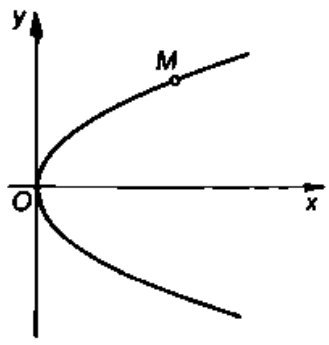


Fig. 41

Equation (28) is the *canonical equation of the parabola*. It is a second-order equation. Thus the parabola is a second-order curve.

Let us now investigate the shape of the parabola using its equation (28). Since equation (28) includes $y$ only in an even degree, the parabola is symmetric with respect to the $x$-axis. Consequently, it is sufficient to consider only its part which lies in the upper half-plane. For this part $y \geqslant 0$ and therefore, solving equation (28) for $y$, we get

$$y = \sqrt{2px}. \tag{29}$$

Relation (29) yields the following statements:

(1) If $x < 0$, then relation (29) yields imaginary values of $y$. Consequently, there is not a single point of the parabola to the left of the $y$-axis. We have noted this fact earlier.

(2) If $x = 0$, then $y = 0$. Thus the origin lies on the parabola and is its leftmost point.

(3) $y$ increases with $x$, and if $x \to +\infty$, then $y \to +\infty$ as well.

Thus the variable point $M(x, y)$, which moves along the parabola with an increase in $x$, emanates from the origin and moves "to the right" and "upwards", the receding of the point $M$ both from the $y$-axis and from the $x$-axis being infinite as $x \to +\infty$.

Making a symmetric mapping about the $x$-axis of the part of the parabola we have considered, we get the whole parabola (Fig. 41) defined by equation (28).

The point 0 is the *vertex of the parabola* and the axis of symmetry (the $x$-axis) is the *axis of the parabola*. The number $p$, i.e. the parameter of the parabola, expresses the distance from the focus to the directrix.

Let us find out how the parameter of the parabola affects its shape. For that purpose, we take some definite value of the abscissa, say, $x = 1$, and use equation (28) to find the corresponding value of the ordinate: $y = \pm\sqrt{2p}$. We get two points $M_1$ $(1, +\sqrt{2p})$ and $M_2$ $(1, -\sqrt{2p})$ on the parabola, which are symmetric about its axis. The distance between them is $2\sqrt{2p}$. Hence we infer that

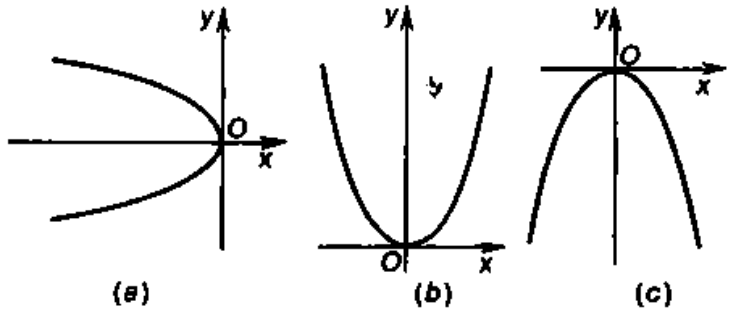



Fig. 42

this distance is the greater, the greater $p$ is. Consequently, the parameter $p$ characterizes the "width" of the domain bounded by the parabola. And this is the geometric meaning of the parameter $p$.

The parabola whose equation is $y^2 = -2px$, $p > 0$, lies to the left of the axis of ordinates (Fig. 42*a*). The vertex of this parabola coincides with the origin, and the $x$-axis is the axis of symmetry.

The equation $x^2 = 2py$, $p > 0$, is the equation of the parabola whose vertex coincides with the origin, and the $y$-axis is the axis of symmetry (Fig. 42*b*). This parabola lies above the abscissa axis. The equation $x^2 = -2py$, $p > 0$, defines the parabola which lies below the $x$-axis, with its vertex at the origin (Fig. 42*c*).

### 1.3.8. The General Equation of a Second-Order Curve

One of the most important problems of analytic geometry is investigating the general equation of a second-order curve and reducing it to the simplest (canonical) form.

The general equation of a second-order curve has the following form:

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F = 0, \qquad (1)$$

where the coefficients $A$,* $2B$, $C$, $2D$, $2E$, $F$* are any numbers and, in addition, the numbers $A$, $B$, $C$ are simultaneously nonzero, i.e. $A^2 + B^2 + C^2 \neq 0$.

**1. Reduction of the general equation of a second-order curve to the canonical form.**

**Lemma 3.1.** *Assume that equation* (1) *is defined in the rectangular system of coordinates $Oxy$ and $AC - B^2 \neq 0$. Then, by means of a parallel displacement and a subsequent rotation of the coordinate axes, equation* (1) *can be reduced to the form*

$$A'x''^2 + C'y''^2 + F' = 0, \tag{2}$$

*where $A'$, $C'$, $F'$ are some numbers and $(x'', y'')$ are the coordinates of the point in the new system of coordinates.*

*Proof.* Assume that we have got a rectangular system of coordinates $O'x'y'$ by a parallel displacement of the $x$-axis and $y$-axis and transferred the origin to the point $O'$ $(x_0, y_0)$. Then the old coordinates $(x, y)$ are related to the new ones $(x', y')$ by the formulas

$$x = x' + x_0, \quad y = y' + y_0$$

(see formulas (1) in 1.3.4). In the new coordinates, equation (1) assumes the form

$$Ax'^2 + 2Bx'y' + Cy'^2 + 2D'x' + 2E'y' + F' = 0, \tag{3}$$

where

$$D' = Ax_0 + By_0 + D, \quad E' = Bx_0 + Cy_0 + E,$$
$$F' = Ax_0^2 + 2Bx_0y_0 + Cy_0^2 + 2Dx_0 + 2Ey_0 + F.$$

In equation (3) the coefficients $D'$ and $E'$ vanish if we choose the coordinates of the point $(x_0, y_0)$ so that the equalities

$$\begin{cases} Ax_0 + By_0 + D = 0, \\ Bx_0 + Cy_0 + E = 0 \end{cases} \tag{4}$$

are satisfied. Since $AC - B^2 \neq 0$, system (4) has a unique solution with respect to $x_0$ and $y_0$.

If the pair of numbers $x_0$, $y_0$ is a solution of system (4), then equation (3) can be written in the form

$$Ax'^2 + 2Bx'y' + Cy'^2 + F' = 0. \tag{5}$$

Assume now that the rectangular system of coordinates $O'x''y''$ results from a rotation of the system $O'x'y'$ through an angle $\alpha$. Then the coordinates $x'$ and $y'$ are related to the coordinates $x''$ and $y''$ by the formulas

$$x' = x''\cos\alpha - y''\sin\alpha, \quad y' = x''\sin\alpha + y''\cos\alpha$$

* To make the transformations of equations (1) more convenient, we designated the coefficients in $xy$, $x$, $y$ as $2B$, $2D$, $2E$ respectively.

(see relations (3) in 1.3.4). In the system of coordinates $O'x''y''$, equation (5) assumes the form

$$A'x''^2 + 2B'x''y'' + C'y''^2 + F' = 0, \tag{6}$$

where

$$A' = A\cos^2\alpha + 2B\cos\alpha\sin\alpha + C\sin^2\alpha,$$

$$B' = -A\sin\alpha\cos\alpha + B(\cos^2\alpha - \sin^2\alpha) + C\sin\alpha\cos\alpha,$$

$$C' = A\sin^2\alpha - 2B\cos\alpha\sin\alpha + C\cos^2\alpha.$$

We choose the angle $\alpha$ such that the coefficient $B'$ in equation (6) vanishes. This requirement leads to an equation $2B\cos 2\alpha = (A - C)\sin 2\alpha$ with respect to $\alpha$. If $A = C$, then $\cos 2\alpha = 0$ and we can set $\alpha = \pi/4$. Now if $A \neq C$, then we choose $\alpha = (1/2) \times \arctan(2B/A - C)$ and equation (6) assumes the form

$$A'x''^2 + C'y''^2 + F' = 0,$$

i.e. we have got equation (2). ■

**Remark.** Equations (4) are known as the *equations of the centre of a second-order curve* and the point $(x_0, y_0)$, where $x_0$ and $y_0$ is a solution of system (4), is the *centre of that curve*. Note that the necessary and sufficient condition for the existence of a unique solution of system (4) is the difference from zero of the number $AC - B^2$, which is the determinant of the system (see 1.10.2).

**2. The invariance of the expression $AC - B^2$. Classification of the second-order curves.** It follows from the proof of Lemma 3.1 that under the parallel displacement of the coordinate axes the coefficients $A$, $B$, $C$ in the leading terms of equation (1) do not change. They change when the axes of coordinates are rotated. However, the expression $AC - B^2$ remains unchanged both under a displacement and under a rotation of the axes, i.e. it does not depend on the transformation of coordinates. Indeed, in the case of a parallel displacement this fact is obvious (see equations (1) and (5)); let us verify it for the case of a rotation of the coordinates. For that purpose, we make use of the expressions for the coefficients $A'$, $B'$, $C'$ of equation (6). We have

$$\begin{aligned} A'C' - B'^2 &= (A\cos^2\alpha + 2B\sin\alpha\cos\alpha + C\sin^2\alpha) \\ &\quad \times (A\sin^2\alpha - 2B\sin\alpha\cos\alpha + C\cos^2\alpha) \\ &\quad - [(C - A)\sin\alpha\cos\alpha + B(\cos^2\alpha - \sin^2\alpha)]^2. \end{aligned}$$

Removing the brackets and collecting terms, we obtain

$$\begin{aligned} A'C' - B'^2 &= AC(\cos^2\alpha + \sin^2\alpha)^2 - B^2(\cos^2\alpha + \sin^2\alpha)^2 \\ &= AC - B^2, \end{aligned}$$

and this is what we wished to prove.

The quantity $AC - B^2$ is the *invariant of the general equation of a second-order curve.* It plays an important part in the investigation of second-order curves.

Depending on the sign of the quantity $AC - B^2$, the second-order curves are classified into three types:

(1) elliptic if $AC - B^2 > 0$;

(2) hyperbolic if $AC - B^2 < 0$;

(3) parabolic if $AC - B^2 = 0$.

Let us consider curves of these types.

(1) **The elliptic type.** Since $AC - B^2 > 0$, it follows, according to Lemma 3.1, that the general equation of the second-order curve can be reduced to the form (to make the notation more convenient, we omit the primes in the coefficients and coordinates)

$$Ax^2 + Cy^2 + F = 0.$$

The following cases are possible here:

(a) $A > 0$, $C > 0$ (the case $A < 0$, $C < 0$ reduces to the case $A > 0$, $C > 0$ when the equation is multiplied by $-1$), $F < 0$. We transfer $F$ to the right-hand side of the equation and divide the equation by it. The equation assumes the form

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

where $a^2 = -F/A$, $b^2 = -F/C$. Comparing the equation obtained with that of the ellipse (see equation (7) in 1.3.7), we infer that it is the canonical equation of the ellipse.

(b) $A > 0$, $C > 0$, $F > 0$. Then, by analogy with the preceding case, we can reduce the equation to the form

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = -1.$$

This equation is not satisfied by the coordinates of any point of the plane. It is known as the *equation of the imaginary ellipse.*

(c) $A > 0$, $C > 0$, $F = 0$. The equation has the form ($a^2 = A$, $c^2 = C$)

$$a^2x^2 + c^2y^2 = 0.$$

It is satisfied by the coordinates of only one point $x = 0$, $y = 0$. This is the *equation of a pair of imaginary intersecting straight lines.*

(2) **The hyperbolic type.** Since $AC - B^2 < 0$, it follows, according to Lemma 3.1, that the general equation of the second-order curve reduces to the form

$$Ax^2 + Cy^2 + F = 0.$$

The following cases are possible here:

(a) $A > 0$, $C < 0$ (the case $A < 0$, $C > 0$ reduces to the case $A > 0$, $C < 0$ when the equation is multiplied by $-1$), $F \neq 0$.

Let $F < 0$, for instance. We transfer $F$ to the right-hand side of the equation and divide the equation by it. The equation assumes the form

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1,$$

where $a^2 = -F/A$, $b^2 = F/C$. Comparing it with the equation of the hyperbola (see equation (15) in 1.3.7), we infer that the equation obtained is the canonical equation of the hyperbola.

(b) $A > 0$, $C < 0$, $F = 0$. The equation assumes the form ($a^2 = A$, $c^2 = -C$)

$$a^2x^2 - c^2y^2 = 0, \quad \text{or} \quad (ax - cy)(ax + cy) = 0.$$

The last equation is satisfied only by the coordinates of the points of the plane which lie on the straight lines $(ax - cy) = 0$ and $(ax + cy) = 0$ which meet at the origin, and thus we have a *pair of intersecting straight lines.*

(3) **The parabolic type.** If $AC - B^2 = 0$, we can reduce the general equation of the second-order curve to the form

$$Ax^2 + Cy^2 + 2Ey + 2Dx + F = 0 \tag{7}$$

by rotating the axes of coordinates through the same angle $\alpha$ as in Lemma 3.1. Here $AC = 0$ and, consequently, one of the coefficients, $A$ or $C$, is zero.

Let $A = 0$, $C \neq 0$. We represent equation (7) as

$$C\left[y^2 + \frac{2E}{C}y + \left(\frac{E}{C}\right)^2\right] + 2Dx + F - \frac{E^2}{C} = 0,$$

or

$$C\left(y + \frac{E}{C}\right)^2 + 2Dx + F^* = 0,$$

where $F^* = F - E^2/C$. We displace the origin parallel to the $y$-axis to the point $(0, -E/C)$, i.e. pass to new coordinates using the formulas $x' = x$, $y' = y + E/C$. We get an equation

$$Cy'^2 + 2Dx' + F^* = 0.$$

The following cases are possible here:

(a) $D \neq 0$. We write the equation in the form

$$Cy'^2 + 2D\left(x' + \frac{F^*}{2D}\right) = 0.$$

Now we displace the origin parallel to the $x'$-axis to the point $(-F^*/(2D), 0)$, i.e. pass to new coordinates using the formulas $x'' = x' + F^*/(2D)$, $y'' = y'$. We get an equation

$$Cy''^2 + 2Dx'' = 0, \text{ or } y''^2 = 2px'',$$

where $p = -D/C$. Comparing the last equation with the equation of the parabola (see equation (28) in 1.3.7), we infer that it is the canonical equation of the parabola.

(b) $D = 0$. The equation has the form

$$Cy'^2 + F^* = 0.$$

If $C$ and $F^*$ are of different signs, then, setting $|F^*/C| = a^2$, we can write the equation in the form

$$(y' - a)(y' + a) = 0.$$

This equation defines a *pair of parallel straight lines.*

If $C$ and $F^*$ are of the same sign, then the equation assumes the form

$$y'^2 + a^2 = 0.$$

This equation is not satisfied by the coordinates of any point of the plane. It is the *equation of a pair of imaginary parallel straight lines.*

Finally, if $F^* = 0$, then the equation assumes the form

$$y'^2 = 0$$

and defines the $O'x'$ axis. We can consider this equation to be a limiting case for $F^* \to 0$, i.e. to be the *equation of a pair of coincident straight lines.*

Now we finish with the investigation of the general equation of the second-order curve and can formulate the results obtained as a theorem.

**Theorem 3.7.** *Let the general equation of the second-order curve*

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F = 0$$

*be given in a rectangular system of coordinates. Then there is a rectangular system of coordinates in which this equation assumes one of the following nine canonical forms*: (1) $(x^2/a^2) + (y^2/b^2) = 1$ (*the ellipse*); (2) $(x^2/a^2) + (y^2/b^2) = -1$ (*the imaginary ellipse*); (3) $a^2x^2 + c^2y^2 = 0$ (*a pair of imaginary intersecting straight lines*); (4) $(x^2/a^2) - (y^2/b^2) = 1$ (*the hyperbola*); (5) $a^2x^2 - c^2y^2 = 0$ (*a pair of intersecting straight lines*); (6) $y^2 = 2px$ (*the parabola*); (7) $y^2 - a^2 = 0$ (*a pair of parallel straight lines*); (8) $y^2 + a^2 = 0$ (*a pair of imaginary parallel straight lines*); (9) $y^2 = 0$ (*a pair of coincident straight lines*).

## Chapter 4

## FUNCTIONS OF ONE VARIABLE

We begin now studying the concept of a function, which is one of the most important concepts in analysis. In this chapter we shall introduce the concept of a limit of a function and the concept of the continuity of a function.

### 1.4.1. The Function

**1. Definition of the function.**

**Definition.** *Let $X$ and $Y$ be some number sets. A function is the set $f$ of ordered pairs of numbers $(x, y)$* such that $x \in X$, $y \in Y$ and each $x$ enters into one and only one pair of this set and each $y$ enters into at least one pair. Then we say that the number $x$ is put into correspondence with a number $y$ and write $y = f(x)$. The number $y$ is the value of the function $f$ at the point $x$. The variable $y$ is a dependent variable and the variable $x$ is an independent variable (or argument), the set $X$ is the domain of definition (or existence) of the function and the set $Y$ is the set of values of the function (or its range).*

Other letters, besides $f$, are used to denote a function, say, $y = y(x)$, $y = g(x)$, $y = \varphi(x)$, $y = A(x)$, $y = F(x)$ etc. Dependent and independent variables can also be denoted by other letters. Sometimes a dependent variable is also called a function.

Alongside the term "function" we sometimes use the equivalent term "mapping" and instead of writing $y = f(x)$ we write $f: x \to y$ and say that the mapping $f$ transforms the number $x$ into the number $y$, or, what is the same, the number $y$ is the image of the number $x$ under the mapping $f$.

The notation $y = f(x)$ is usually more convenient in calculations than the notation $f: x \to y$. For instance, the notation $f(x) = x^2$ is considerably more convenient to use in analytical transformations than the notation $f: x \to x^2$.

A function whose all values are equal is said to be *constant*. It is often designated as $C$.

When a function $f(x)$ is defined on a set $X$, we say that it is *bounded from above* (*from below*) on this set if there is a number $M$ ($m$) such that the inequality $f(x) \leqslant M$ ($f(x) \geqslant m$) is satisfied for any $x \in X$. A function which is bounded both from above and from below on the set $X$ is said to be *bounded* on this set. The condition for the boundedness of the function $f(x)$ can be written as follows: there is a number $M > 0$ such that the inequality $|f(x)| \leqslant M$ is satisfied for any $x \in X$. For instance, the function $f(x) = \sin x$ is bounded on the entire number line since $|\sin x| \leqslant 1$ for any $x$, and the function $f(x) = 1/x$ is not bounded from above on the interval $(0, 1)$ since there is no number $M$ such that the inequality $1/x \leqslant M$ would be satisfied for any $x \in (0, 1)$.

On the plane, a function is represented as a graph which is the set of points $(x, y)$ whose coordinates are connected by the relation $y = f(x)$ called the *equation of the graph*.

* Recall that the pair of numbers $x$ and $y$ is said to be *ordered* if it is indicated which of these numbers is the first and which is the second. An ordered pair of numbers is written as $(x, y)$, where $x$ is the first number and $y$ is the second.

The graph of a function may be a solid line (a curve or a straight line) or may consist of separate points, as, for instance, the graph of the function $y = n!$ in Fig. 45.

Note that not every line is the graph of a function. For instance, the circle $x^2 + y^2 = 1$ is not a graph of a function since every

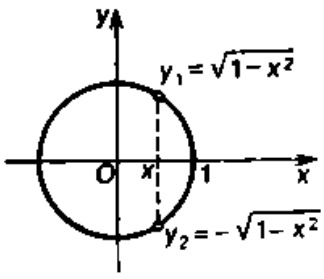



Fig. 43

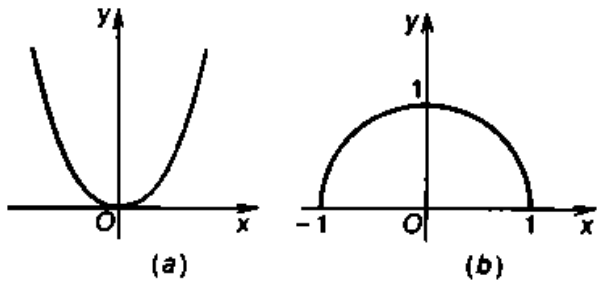



Fig. 44

$x \in (-1, 1)$ belongs not to one but to two pairs of numbers $(x, y)$ of this set with different values of $y$: $y_1 = \sqrt{1 - x^2}$ and $y_2 = -\sqrt{1 - x^2}$, and this contradicts the requirement of single-valuedness in the definition of a function (Fig. 43). However, the part of the circle which lies in the lower half-plane is the graph of the function $y = -\sqrt{1 - x^2}$ and its other part which lies in the upper half-plane is the graph of the function $y = \sqrt{1 - x^2}$.

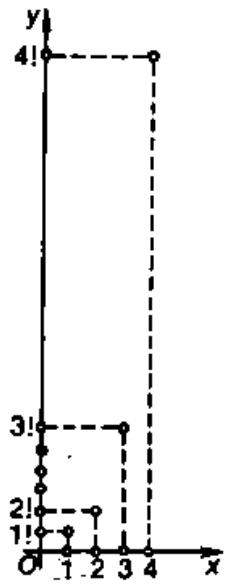



Fig. 45

**2. Methods of representing functions.** To define a function $f$ is to show how, using every value of the argument $x$, we can find the corresponding value of the function $f(x)$. There are three principal methods of representing functions: *analytical*, *tabular* and *graphical*.

**(1) Analytical method.** This method consists in defining the relationship between the variables by means of a formula which indicates the actions that must be performed to obtain the value of the function corresponding to the given value of the argument.

Let us consider some examples.

1. The formula $y = x^2$ defines a function whose domain of definition is the number line $(-\infty, +\infty)$ and the range is a half-line $[0, +\infty)$ (Fig. 44*a*).

2. The formula $y = \sqrt{1 - x^2}$ defines a function whose domain of definition is the closed interval $[-1, 1]$ and the range is the closed interval $[0, 1]$ (Fig. 44*b*).

3. The formula $y = n!$ associates every natural number (i.e. a pos-

itive integer) $n$ with the number $y = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$. For example, if $n = 3$, then $y = 3! = 6$. Thus the formula $y = n!$ defines a function whose domain of definition is $\{1, 2, 3, \ldots, n, \ldots\}$ and the range is $\{1!, 2!, 3!, \ldots, n!, \ldots\}$ (Fig. 45).

4. $$y = \operatorname{sgn} x^* = \begin{cases} +1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -1 & \text{if } x < 0. \end{cases}$$

This function is defined by several formulas. It is defined on the entire number line $(-\infty, +\infty)$ and the set of its values consists of three numbers: $-1$, $0$, $+1$ (Fig. 46).

5. Dirichlet's** function

$$y = \begin{cases} 1 \text{ if } x \text{ is a rational number,} \\ 0 \text{ if } x \text{ is an irrational number.} \end{cases}$$

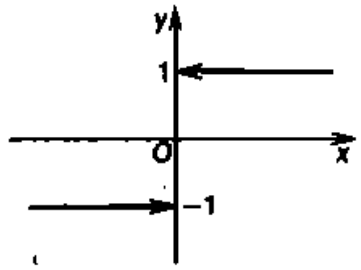


Fig. 46

This function is defined throughout the number line $(-\infty, +\infty)$, and the set of its values consists of two numbers 0 and 1.

Note that it is impossible to represent Dirichlet's function by graphical means.

**(2) Tabular method.** We present the following table:

| $x$ | 0 | 0.1 | 0.2 | 3 | 0.6 | 4 | 0.8 | 1.5 | 2 |
|---|---|---|---|---|---|---|---|---|---|
| $y$ | −1 | 10 | 1 | −2 | −8 | 0.5 | −2 | 5 | 7 |

We put every $x$ written in the first row of the table into correspondence with the number $y$ written in the second row under that number $x$ and say that the function obtained is specified by the table. The domain of definition of this function is a set consisting of nine numbers $x$ enumerated in the first row of the table and its range is a set consisting of nine numbers $y$ enumerated in the second row.

With the aid of a table, we can specify a function only for a finite number of values of the argument. Tables are often used to specify functions. Thus we are familiar with tables of trigonometric functions, tables of logarithms and many other tables. A timetable of trains which defines the location of a train at separate time moments may serve as an example of the tabular representation of functions.

* The term sgn comes from the Latin word "signum" meaning sign.

** Dirichlet, Peter Gustav Lejeune (1805-1859), a German mathematician.

**(3) Graphical method.** The graphical method of representation of functions is usually used in physical measurements when the correspondence between the variables $x$ and $y$ is defined by a graph. In many cases graphs of this kind are drawn by means of recording instruments. Thus, for instance, to measure atmospheric pressure at different altitudes, use is made of a special recorder, a barograph, which records the variation of atmospheric pressure with altitude on a moving sheet of graph paper.

**3. Classification of functions.** The constant function $f(x) = C$, $C = \text{const}$, the power function $x^\alpha$ ($\alpha$ is any number), the exponential function $a^x$ $(0 < a \neq 1)$, the logarithmic function $\log_a x$ $(0 < a \neq 1)$, the trigonometric functions $\sin x$, $\cos x$, $\tan x$, $\cot x$ and the inverse trigonometric functions $\arcsin x$, $\arccos x$, $\arctan x$, $\operatorname{arccot} x$ are all the *simplest* (or *basic*) *elementary functions.*

All functions which can be obtained by means of a finite number of arithmetic operations involving the simplest elementary functions as well as by a superposition of these functions form a class of *elementary functions.* The following functions can serve as examples of elementary functions: $f(x) = |x|$ $(|x| = \sqrt{x^2})$, $f(x) = \log^2 \arctan 2^{\sqrt{x}} + \sin 3x$, $f(x) = \ln|\sin 3x| - e^{\arctan\sqrt{x}}$ etc.

We have the following classification of elementary functions.

(1) A function of the form

$$P(x) = a_0x^m + a_1x^{m-1} + \ldots + a_{m-1}x + a_m,$$

where $m \geqslant 0$ is an integer, $a_0, a_1, \ldots, a_m$ are any numbers which are coefficients $(a_0 \neq 0)$, is said to be an *integral rational function* or an *algebraic polynomial of degree m.* A first-degree polynomial is also called a *linear function.*

(2) A function, which is the ratio of two integral rational functions

$$R(x) = \frac{a_0x^m + a_1x^{m-1} + \ldots + a_{m-1}x + a_m}{b_0x^n + b_1x^{n-1} + \ldots + b_{n-1}x + b_n},$$

is known as a *fractional rational function.*

Integral rational functions and fractional rational functions taken together form a class of *rational functions.*

(3) A function, which is obtained as a result of a finite number of superpositions and four arithmetic operations involving power functions with both integral and fractional exponents and which is not rational, is an *irrational function.* For example, $f(x) = \sqrt{x}$, $f(x) = x + \sqrt{x}$, $f(x) = \sqrt{(5x^2 + 4x - 7)/(3x^2 - 8x + 4)} + (\sqrt{x} + x)^3$ etc are irrational functions.

(4) Every function which is not rational or irrational is a *transcendental function.* These are, for instance, the functions $f(x) = \sin x$, $f(x) = \sin x + x$ etc.

### 1.4.2. The Limit of a Function

**1. The limit of a function for $x \to x_0$.** Assume that the function $f(x)$ is defined on a set $X$ and let the point $x_0 \in X$ or $x_0 \notin X$. We take from $X$ a sequence of points different from $x_0$, i.e.

$$x_1, x_2, x_3, \ldots, x_n, \ldots, \tag{1}$$

which converges to $x_0$.* The values of the function at the points of this sequence also form a number sequence

$$f(x_1), f(x_2), f(x_3), \ldots, f(x_n), \ldots \tag{2}$$

and we can raise the question as to whether it possesses a limit.

**Definition 1.** *The number $A$ is the limit of the function $f(x)$ at the point $x = x_0$ (or as $x \to x_0$) if for any, converging to $x_0$, sequence*

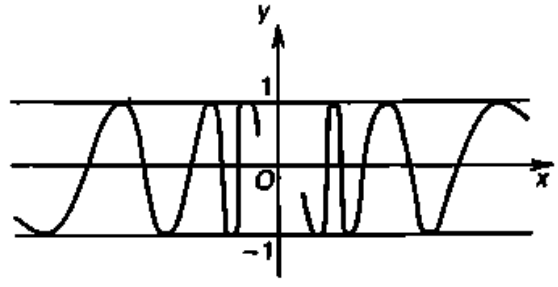



Fig. 47

(1) *of the values of the argument $x$, which differ from $x_0$, there is a sequence* (2) *of the values of the function, which converges to the number $A$.*

Using symbols, we can write this as $\lim\limits_{x \to x_0} f(x) = A$.

The function $f(x)$ can have only one limit at the point $x_0$. This follows from the fact that the sequence $\{f(x_n)\}$ has only one limit.

Let us consider some examples.

1. The function $f(x) = C = \text{const}$ has a limit at the point $x_0$ of the number line. Indeed, if (1) is any sequence converging to $x_0$, then sequence (2) has the form $C, C, \ldots, C, \ldots$, i.e. $f(x_n) = C$. Hence we infer that $f(x_n) \to C$ as $n \to \infty$ or $\lim\limits_{x \to x_0} f(x) = C$.

2. At any point $x_0$ of the number line, the function $f(x) = x$ has a limit equal to $x_0$. In this case sequences (1) and (2) are identical, i.e. $f(x_n) = x_n$. Consequently, if $x_n \to x_0$, then $f(x_n) \to x_0$ as $n \to \infty$ or $\lim\limits_{x \to x_0} f(x) = \lim\limits_{x \to x_0} x = f(x_0) = x_0$.

3. The function $f(x) = \sin(1/x)$ (Fig. 47) defined for all $x \neq 0$ has no limit at the point $x = 0$. Indeed, we take two sequences of

* We assume that such a sequence does exist.

the values of the argument $x$: $1/\pi, 1/(2\pi), 1/(3\pi), \ldots, 1/(n\pi), \ldots$ and $2/\pi, 2/(5\pi), 2/(9\pi), \ldots, 2/[(4n-3)\pi], \ldots$, which converge to zero. The sequences of the values of the function corresponding to them are $f\left(\frac{1}{\pi}\right), f\left(\frac{1}{2\pi}\right), f\left(\frac{1}{3\pi}\right), \ldots, f\left(\frac{1}{n\pi}\right), \ldots$ and $f\left(\frac{2}{\pi}\right), f\left(\frac{2}{5\pi}\right), f\left(\frac{2}{9\pi}\right), \ldots, f\left(\frac{2}{(4n-3)\pi}\right), \ldots$. Since $f\left(\frac{1}{n\pi}\right) = \sin n\pi = 0$ and $f\left(\frac{2}{(4n-3)\pi}\right) = \sin\frac{(4n-3)\pi}{2} = 1$ for any $n$, we have $\lim\limits_{n\to\infty} f(x_n) = \lim\limits_{n\to\infty} \sin n\pi = 0$ for the first sequence and $\lim\limits_{n\to\infty} f(x_n) = \lim\limits_{n\to\infty} \sin\frac{(4n-3)\pi}{2} = 1$ for the second sequence. Thus, for two sequences of the values of the argument $x$ converging to zero, the corresponding sequences of the values of the function have different limits. And this means, according to the definition of the limit of a function, that $\lim\limits_{x\to 0} f(x)$ does not exist.

4. The function $f(x) = \frac{x^2+x-1}{x-1}$ has a limit equal to 1 at the point $x = 0$. Indeed, we take any sequence of the values of the argument $x$, which converges to zero, i.e. $\lim\limits_{n\to\infty} x_n = 0$, and $x_n \neq 0$, and then, by virtue of Theorems 2.7-2.9, we have

$$\lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} \frac{x_n^2+x_n-1}{x_n-1} = \frac{(\lim\limits_{n\to\infty} x_n)^2 + \lim\limits_{n\to\infty} x_n - 1}{\lim\limits_{n\to\infty} x_n - 1} = 1.$$

Thus there is $\lim\limits_{n\to\infty} f(x_n) = 1$ and, since it does not depend on the choice of the sequence $\{x_n\}$ converging to zero, we infer, by virtue of the definition of the limit of a function, that $\lim\limits_{x\to 0} f(x) = 1$.

5. Dirichlet's function whose values are equal to unity at rational points and to zero at irrational points has no limit at any point $x_0$ of the number line. Indeed, for a sequence of the rational values of the argument, which converges to the point $x_0$, the limit of the corresponding sequence of the values of the function is unity, and for a sequence of the irrational values of the argument, which converges to the point $x_0$, the limit of the corresponding sequence of the values of the function is zero.

There is another definition of the limit of a function.

**Definition 2.** *The number $A$ is the limit of the function $f(x)$ at the point $x = x_0$ if for any number $\varepsilon > 0$ there is a number $\delta > 0$ such that the inequality $|f(x) - A| < \varepsilon$ holds true for all $x \in X$, $x \neq x_0$, which satisfy the inequality $|x - x_0| < \delta$.*

Using logical symbols, we can write Definition 2 in the form

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall x \in X,\ x \neq x_0,\ |x - x_0| < \delta):\ |f(x) - A| < \varepsilon.$$

Note that we can write the inequalities $x \neq x_0$ and $|x - x_0| < \delta$ as $0 < |x - x_0| < \delta$.

The first definition is based on the concept of a limit of a number sequence and therefore it is often called a definition "in the language of sequences". The second definition is said to be a definition "in the $\varepsilon$-$\delta$ language".

**Theorem 4.1.** *The first and the second definition of the limit of a function are equivalent.*

*Proof.* (1) Let $A$ be the limit of $f(x)$ at the point $x_0$ according to the first definition. We shall show that $A$ is a limit of $f(x)$ according to the second definition. We assume the contrary, i.e. $A$ is not a limit of this function according to the second definition. This means that not for any $\varepsilon > 0$ we can indicate $\delta > 0$ such that the inequality $0 < |x - x_0| < \delta$ would imply the inequality $|f(x) - A| < \varepsilon$, i.e. there is $\varepsilon = \varepsilon_0 > 0$ such that whatever $\delta > 0$ we take, there is at least one point $x \neq x_0$ such that $|x - x_0| < \delta$ but $|f(x) - A| \geqslant \varepsilon_0$. We shall take as $\delta$ the following numbers in succession:

$$1,\ \frac{1}{2},\ \frac{1}{3},\ \ldots,\ \frac{1}{n},\ \ldots.$$

Then

for $\delta = 1$ there is $x_1 \neq x_0$ in $X$ such that $|x_1 - x_0| < 1$ and $|f(x_1) - A| \geqslant \varepsilon_0$,

for $\delta = 1/2$ there is $x_2 \neq x_0$ in $X$ such that $|x_2 - x_0| < 1/2$ and $|f(x_2) - A| \geqslant \varepsilon_0$,

for $\delta = 1/3$ there is $x_3 \neq x_0$ in $X$ such that $|x_3 - x_0| < 1/3$ and $|f(x_3) - A| \geqslant \varepsilon_0$,

. . . . . .

for $\delta = 1/n$ there is $x_n \neq x_0$ in $X$ such that $|x_n - x_0| < 1/n$ and $|f(x_n) - A| \geqslant \varepsilon_0$.

. . . . . .

As a result, we get a sequence of points different from $x_0$:

$$x_1,\ x_2,\ x_3,\ \ldots,\ x_n,\ \ldots,$$

which converges to the point $x_0$ since $|x_n - x_0| < 1/n \to 0$ as $n \to \infty$. Therefore, in accordance with the first definition of the limit of a function, the corresponding sequence $\{f(x_n)\}$ of the values of the function converges to the number $A$. Consequently, for $\varepsilon_0$ there is a number $N$ such that the inequality $|f(x_n) - A| < \varepsilon_0$ is satisfied for all $n > N$. But this cannot occur since the inequality

$|f(x_n) - A| \geqslant \varepsilon_0$ is satisfied for all $x_n$. The contradiction obtained proves that the number $A$ is the limit of the function $f(x)$ at the point $x_0$ in accordance with the second definition.

(2) Assume now that $A$ is the limit of $f(x)$ at the point $x_0$ in accordance with the second definition. This means that for any $\varepsilon > 0$ there is $\delta > 0$ such that the inequality $0 < |x - x_0| < \delta$ implies the inequality $|f(x) - A| < \varepsilon$. We shall show that $A$ is a limit of $f(x)$ in accordance with the first definition. We take an arbitrary sequence of points

$$x_1, x_2, x_3, \ldots, x_n, \ldots,$$

which converges to the point $x_0$ $(x_n \neq x_0)$. Then, for the given value of $\delta > 0$, which corresponds to $\varepsilon$ in accordance with the second definition, there is an $N$ such that the inequalities $0 < |x_n - x_0| < \delta$ are satisfied for $n > N$. But at the same time, by virtue of the second definition, the inequality $|f(x_n) - A| < \varepsilon$ is also satisfied. And since we have chosen $\varepsilon$ arbitrarily, this means that $f(x_n) \to A$ for any sequence $\{x_n\}$, which converges to the point $x_0$ $(x_n \neq x_0)$, i.e. the number $A$ is the limit of $f(x)$ at the point $x_0$ in accordance with the first definition. ■

We have thus established the equivalence of the two definitions of the limit of a function and can now use any of them according to which of the definitions is more convenient in solving a certain problem.

Note that the definition of the limit of a function "in the language of sequences" is also called the definition of the limit of a function according to Heine* and the definition of the limit of a function "in the $\varepsilon$-$\delta$ language" is called the definition of the limit of a function according to Cauchy**.

**2. The limit of a function for $x \to x_0-$ and for $x \to x_0+$.** In what follows, we shall use the concepts of one-sided limits of a function, which are defined as follows.

**Definition 3.** *The number $A$ is the right-hand (left-hand) limit of the function $f(x)$ at the point $x_0$ if for any sequence* (1) *converging to $x_0$, whose elements $x_n$ are greater (smaller) than $x_0$, the corresponding sequence* (2) *converges to $A$.*

The symbolic notation is $\lim\limits_{x \to x_0+} f(x) = A$ $\left(\lim\limits_{x \to x_0-} f(x) = A\right)$. The function $f(x) = \operatorname{sgn} x$*** can serve as an example. This function has both a right-hand limit $\lim\limits_{x \to 0+} \operatorname{sgn} x = 1$ and a left-hand limit $\lim\limits_{x \to 0-} \operatorname{sgn} x = -1$ at the point $x = 0$. Indeed, if (1)

* Heine, Heinrich Eduard (1821-1881), a German mathematician.

** Cauchy, Augustin Louis (1789-1857), a French mathematician.

*** You can find the definition of the function $f(x) = \operatorname{sgn} x$ in 1.4.1, Item 2.

is any sequence of the values of the argument of this function converging to zero, whose elements $x_n$ are greater than zero ($x_n > 0$), then $\operatorname{sgn} x_n = 1$ and $\lim\limits_{n\to\infty} \operatorname{sgn} x_n = 1$. Consequently, $\lim\limits_{x\to 0+} \operatorname{sgn} x = 1$.

By analogy we find that $\lim\limits_{x\to 0-} \operatorname{sgn} x = -1$.

We can give an equivalent definition of one-sided limits of a function "in the $\varepsilon$-$\delta$ language": *the number $A$ is the right-hand (left-hand) limit of the function $f(x)$ at the point $x_0$ if for any $\varepsilon > 0$ there is $\delta > 0$ such that the inequality $|f(x) - A| < \varepsilon$ holds true for all $x$ which satisfy the inequalities $x_0 < x < x_0 + \delta$ ($x_0 - \delta < x < x_0$).* The symbolic notation is

$$(\forall \varepsilon > 0)\ (\exists \delta > 0)\ (\forall x,\quad x_0 < x < x_0 + \delta\ (x_0 - \delta < x < x_0)):$$
$$|f(x) - A| < \varepsilon.$$

The following theorem establishes a relationship between one-sided limits and the limit of a function.

**Theorem 4.2.** *The function $f(x)$ has a limit at the point $x_0$ if and only if both the right-hand and left-hand limits exist at that point and they are equal. In this case the limit of the function is equal to one-sided limits.*

*Proof.* Let $\lim\limits_{x\to x_0-} f(x) = \lim\limits_{x\to x_0+} f(x) = A$. Then, in accordance with the definition of the limit of a function on the left and on the right, for any $\varepsilon > 0$ there are numbers $\delta_1 > 0$ and $\delta_2 > 0$ such that the inequality $|f(x) - A| < \varepsilon$ holds true for all $x$ which satisfy the inequalities $x_0 - \delta_1 < x < x_0$ and for all $x$ which satisfy the inequalities $x_0 < x < x_0 + \delta_2$. We take $\delta = \min\{\delta_1, \delta_2\}$. Then the inequality $|f(x) - A| < \varepsilon$ will hold true for all $x$ which satisfy the inequalities $|x - x_0| < \delta$, $x \neq x_0$, and this means, in accordance with Definition 2, that $\lim\limits_{x\to x_0} f(x) = A$.

Conversely, let $\lim\limits_{x\to x_0} f(x) = A$. Then, in accordance with the definition of the limit of a function at the point $x_0$, for any $\varepsilon > 0$ there is a number $\delta > 0$ such that the inequality $|f(x) - A| < \varepsilon$ holds true for all $x$ which satisfy the inequality $|x - x_0| < \delta$, $x \neq x_0$. Thus the inequality $|f(x) - A| < \varepsilon$ holds true both for $x_0 - \delta < x < x_0$ and for $x_0 < x < x_0 + \delta$, and this, in accordance with the definition of one-sided limits, means that $\lim\limits_{x\to x_0-} f(x) = \lim\limits_{x\to x_0+} f(x) = A$. ■

**3. The limit of a function as $x \to \infty$, as $x \to -\infty$, as $x \to +\infty$.** In addition to the concepts of the limit of a function as $x \to x_0$ and one-sided limits discussed above, there is also a concept of the limit of a function for the argument tending to infinity.

**Definition 4.** *The number $A$ is the limit of the function $f(x)$ as $x \to \infty$*

*if, for any infinitely large sequence* (1) *of the values of the argument, the corresponding sequence* (2) *of the values of the function converges to* $A$.

The symbolic notation is $\lim\limits_{x\to\infty} f(x) = A$.

**Definition 5.** *The number* $A$ *is the limit of the function* $f(x)$ *as* $x \to +\infty$ $(x \to -\infty)$ *if for any infinitely large sequence of the values of the argument, whose elements* $x_n$ *are positive* (*negative*), *the corresponding sequence of the values of the function converges to* $A$.

The symbolic notation is $\lim\limits_{x\to+\infty} f(x) = A$ $\left(\lim\limits_{x\to-\infty} f(x) = A\right)$.

Let us consider an example. Assume that $f(x) = 1/x$. This function has a limit which is equal to zero as $x \to \infty$. Indeed, if $\{x_n\}$ is an infinitely large sequence of the values of the argument, then the corresponding sequence of the values of the function, i.e. $1/x_1, 1/x_2, \ldots, 1/x_3, \ldots$, is infinitely small according to Theorem 2.1, and therefore has a limit equal to zero, i.e. $\lim\limits_{x\to\infty} (1/x) = 0$ (Fig. 48).

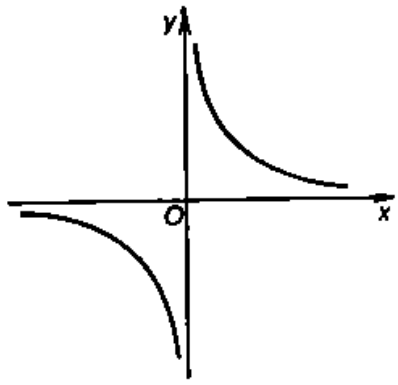


Fig. 48

We have given Definitions 4 and 5 "in the language of sequences". We can give equivalent definitions "in the $\varepsilon$-$\delta$ language" and write them by means of logical symbols. We invite the reader to do that. As an example we shall formulate the definition of the limit of a function as $x \to +\infty$.

**Definition 6.** *The number* $A$ *is the limit of the function* $f(x)$ *as* $x \to +\infty$ *if for any number* $\varepsilon > 0$ *there is a number* $\delta$ *such that the inequality* $|f(x) - A| < \varepsilon$ *holds true for all* $x \in X$ *which satisfy the inequality* $x > \delta$.

### 1.4.3. Theorems on the Limits of Functions

The definition of the limit of a function "in the language of sequences" makes it possible to extend the theorems on the limits of sequences, proved above, to functions. We shall show this on the example of two theorems.

**Theorem 4.3.** *Let the functions* $f(x)$ *and* $g(x)$ *have limits* $B$ *and* $C$ *at the point* $x_0$. *Then the functions* $f(x) \pm g(x)$, $f(x)g(x)$, $\frac{f(x)}{g(x)}$ (*for* $C \neq 0$) *have limits equal to* $B \pm C$, $BC$, $\frac{B}{C}$ *respectively at the point* $x_0$.

*Proof.* Let $\{x_n\}$ $(x_n \neq x_0)$ be an arbitrary sequence of the values of the argument of the functions $f(x)$ and $g(x)$ converging to $x_0$. The corresponding sequences $\{f(x_n)\}$ and $\{g(x_n)\}$ of the values of

these functions have limits $B$ and $C$. But then, by virtue of Theorems 2.7-2.9, the sequences $\{f(x_n) \pm g(x_n)\}$, $\{f(x_n)\, g(x_n)\}$, $\{f(x_n)/g(x_n)\}$ (for $C \neq 0$) have limits equal to $B \pm C$, $BC$, $\frac{B}{C}$ respectively. According to Definition 1 of the limit of a function this means that

$$\lim_{x \to x_0} [f(x) \pm g(x)] = B \pm C, \quad \lim_{x \to x_0} [f(x)\, g(x)] = BC, \quad \lim_{x \to x_0} \frac{f(x)}{g(x)} = \frac{B}{C}. \quad \blacksquare$$

**Theorem 4.4.** *Assume that the functions $f(x)$, $g(x)$, $h(x)$ are defined in a certain neighbourhood of the point $x_0$, except maybe for the point $x_0$ itself, and the functions $f(x)$ and $h(x)$ have a limit equal to $A$ at the point $x_0$, i.e.* $\lim\limits_{x \to x_0} f(x) = \lim\limits_{x \to x_0} h(x) = A$. *Assume, in addition, that the inequalities $f(x) \leqslant g(x) \leqslant h(x)$ are satisfied. Then*

$$\lim_{x \to x_0} g(x) = A.$$

*Proof.* Let $\{x_n\}$ $(x_n \neq x_0)$ be an arbitrary sequence of the values of the argument of the functions $f(x)$ and $h(x)$ converging to $x_0$. The corresponding sequences $\{f(x_n)\}$ and $\{h(x_n)\}$ of the values of these functions have a limit equal to $A$, i.e. $f(x_n) \to A$ and $h(x_n) \to A$ as $n \to \infty$. Using the inequalities given in the hypothesis, we can write

$$f(x_n) \leqslant g(x_n) \leqslant h(x_n).$$

It follows, according to Theorem 2.11, that $g(x_n) \to A$. In accordance with Definition 1 of the limit of a function this means that

$$\lim_{x \to x_0} g(x) = A. \quad \blacksquare$$

**Remark.** Theorems 4.3 and 4.4 are also true in the case when $x_0$ is one of the symbols $\infty$, $+\infty$ or $-\infty$.

### 1.4.4. Two Remarkable Limits

**1. The first remarkable limit.** We shall prove that

$$\lim_{x \to 0} \frac{\sin x}{x} = 1.$$

We shall consider an arc of a circle of radius $R = 1$ with the central angle whose radian measure is $x$ $(0 < x < \pi/2)$ (Fig. 49). Then

$$OA = 1, \ \sin x = MK, \ \tan x = AT. \tag{1}$$

The area of the triangle $OAM$ is evidently smaller than the area of the sector $OAM$ which, in turn, is smaller than the area of the

triangle $OAT$, or, what is the same, $(1/2)\ OA \cdot MK < (1/2) OA \cdot \widehat{AM} < (1/2)\ OA \cdot AT$. Taking relations (1) into account, we can write the last relation in the form $(1/2)\ \sin x < (1/2)x < (1/2)\ \tan\ x$, whence we get

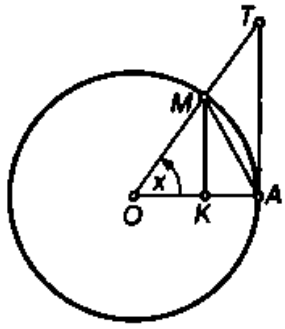


Fig. 49

$$\sin x < x < \tan x. \qquad (2)$$

Dividing these inequalities by $\sin x$, we get $1 > (\sin x)/x > \cos x$, whence we find that $0 < 1 - (\sin x)/x < 1 - \cos x$. Since $\sin (x/2) < 1$, it follows that $\sin^2 (x/2) < \sin (x/2)$. Therefore, taking the first inequality (2) into account, we get $1 - \cos x = 2 \sin^2 (x/2) < 2 \sin (x/2) < 2\ (x/2) = x$ for all $x$ which satisfy the inequalities $0 < x < \pi/2$. Thus $0 < 1 - (\sin x)/x < x$ for $0 < x < \pi/2$.

We take an arbitrary $\varepsilon > 0$ and set $\delta = \min \{\varepsilon, \pi/2\}$. Then the inequality $x < \varepsilon$ will hold true for all $x$ which satisfy the inequalities $0 < x < \delta$ and therefore

$$0 < 1 - \frac{\sin x}{x} < \varepsilon,$$

whence

$$\left| 1 - \frac{\sin x}{x} \right| < \varepsilon.$$

This means that 1 is the right-hand limit of the function $(\sin x)/x$ at the point $x = 0$, i.e. $\lim\limits_{x \to 0+} \frac{\sin x}{x} = 1$. Now we take notice of the fact that the function $f(x) = (\sin x)/x$ is even since $f(-x) = (\sin\ (-x))/-x = (\sin\ x)/x = f(x)$. Therefore the left-hand limit of the function $(\sin x)/x$ is also equal to 1 at the point $x = 0$. By virtue of Theorem 4.2, it follows that $\lim\limits_{x \to 0} \frac{\sin x}{x} = 1$. ■

**Remark.** Using the inequalities $\sin x < x$ and $1 - \cos x < x$ for $0 < x < \pi/2$ we have obtained when discussing the first remarkable limit, we can easily prove that $\lim\limits_{x \to 0} \cos x = 1$ and $\lim\limits_{x \to 0} \sin x = 0$. We invite the reader to do that.

Many other limits can be calculated with the aid of the first remarkable limit.

**Example 1.** Find $\lim\limits_{x \to 0} \frac{1 - \cos x}{x}$.

*Solution.* As $x \to 0$, the denominator of the fraction tends to zero. Therefore Theorem 4.3 is inapplicable here. To find the limit, we

transform the fraction as follows:

$$\lim_{x\to 0}\frac{1-\cos x}{x}=\lim_{x\to 0}\frac{2\sin^2(x/2)}{x}=\lim_{x\to 0}\frac{\sin(x/2)}{x/2}\sin(x/2)$$

$$=\lim_{x\to 0}\frac{\sin(x/2)}{x/2}\lim_{x\to 0}\sin(x/2)=1\cdot 0=0.$$

**Example 2.** Find $\lim\limits_{x\to 0}\frac{\tan x}{x}$.

*Solution.* We have

$$\lim_{x\to 0}\frac{\tan x}{x}=\lim_{x\to 0}\frac{\sin x}{x}\frac{1}{\cos x}=\lim_{x\to 0}\frac{\sin x}{x}\lim_{x\to 0}\frac{1}{\cos x}=1\cdot\frac{1}{1}=1.$$

**Example 3.** Find $\lim\limits_{x\to 0}\frac{5x}{\sin 4x}$.

*Solution.* We have

$$\lim_{x\to 0}\frac{5x}{\sin 4x}=\lim_{x\to 0}\frac{5/4}{(\sin 4x)/(4x)}=\frac{5/4}{\lim\limits_{x\to 0}(\sin 4x)/(4x)}=\frac{5/4}{1}=1.25.$$

**2. The second remarkable limit.** Let us prove that

$$\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x=\mathrm{e}.$$

As is known, $\lim\limits_{n\to\infty}\left(1+\frac{1}{n}\right)^n=\mathrm{e}$ (see 1.2.3, Item 2). Let $x>1$. We set $n=[x]$; then $x=n+\alpha$, where $n$ is a natural number and $\alpha$ satisfies the condition $0\leqslant\alpha<1$. Since $n\leqslant x<n+1$, $1/(n+1)<1/x\leqslant 1/n$, it follows that

$$\left(1+\frac{1}{n+1}\right)^n<\left(1+\frac{1}{x}\right)^x<\left(1+\frac{1}{n}\right)^{n+1}.$$

As $x\to+\infty$ $(n\to\infty)$, we have

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^{n+1}=\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n\lim_{n\to\infty}\left(1+\frac{1}{n}\right)=\mathrm{e}\cdot 1=\mathrm{e}$$

and

$$\lim_{n\to\infty}\left(1+\frac{1}{n+1}\right)^n=\frac{\lim\limits_{n\to\infty}(1+1/(n+1))^{n+1}}{\lim\limits_{n\to\infty}(1+1/(n+1))}=\frac{\mathrm{e}}{1}=\mathrm{e}.$$

From this, in accordance with Theorem 4.4, we get

$$\lim_{x\to+\infty}\left(1+\frac{1}{x}\right)^x=\mathrm{e}.$$

Let now $x<-1$; we set $x=-y$ and then

$$\lim_{x\to-\infty}\left(1+\frac{1}{x}\right)^x=\lim_{y\to+\infty}\left(1-\frac{1}{y}\right)^{-y}=\lim_{y\to+\infty}\left(1+\frac{1}{y-1}\right)^y$$

$$=\lim_{y\to+\infty}\left(1+\frac{1}{y-1}\right)^{y-1}\lim_{y\to+\infty}\left(1+\frac{1}{y-1}\right)=e\cdot 1=e$$

as $x\to-\infty$.

Combining the two cases, we finally get

$$\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x=e. \blacksquare$$

The second remarkable limit has a variety of applications. Many other limits can be found with its aid.

**Example 4.** Find $\lim\limits_{x\to 0}(1+x)^{1/x}$.

*Solution.* We make a change of variable setting $1/x=\alpha$. Then it is evident that $\alpha\to\infty$ as $x\to 0$. Therefore

$$\lim_{x\to 0}(1+x)^{1/x}=\lim_{\alpha\to\infty}\left(1+\frac{1}{\alpha}\right)^{\alpha}=e.$$

**Example 5.** Find $\lim\limits_{x\to\infty}\left(1+\frac{3}{x}\right)^x$.

*Solution.* We set $x=3t$. Then $t\to\infty$ as $x\to\infty$. Consequently,

$$\lim_{x\to\infty}\left(1+\frac{3}{x}\right)^x=\lim_{t\to\infty}\left(1+\frac{1}{t}\right)^{3t}=\lim_{t\to\infty}\left[\left(1+\frac{1}{t}\right)^t\left(1+\frac{1}{t}\right)^t\left(1+\frac{1}{t}\right)^t\right]$$

$$=\lim_{t\to\infty}\left(1+\frac{1}{t}\right)^t\lim_{t\to\infty}\left(1+\frac{1}{t}\right)^t\lim_{t\to\infty}\left(1+\frac{1}{t}\right)^t=e\cdot e\cdot e=e^3.$$

**Example 6.** Find $\lim\limits_{x\to 0}\frac{\log_a(1+x)}{x}$.

*Solution.* To find the limit, we transform the fraction as follows:

$$\lim_{x\to 0}\frac{\log_a(1+x)}{x}=\lim_{x\to 0}\left[\frac{1}{x}\log_a(1+x)\right]$$

$$=\lim_{x\to 0}\log_a(1+x)^{1/x}=\log_a\left[\lim_{x\to 0}(1+x)^{1/x}\right].$$

But $\lim\limits_{x\to 0}(1+x)^{1/x}=e$ (see Example 4). Therefore $\lim\limits_{x\to 0}\frac{\log_a(1+x)}{x}=\log_a e$. In particular, $\lim\limits_{x\to 0}\frac{\ln(1+x)}{x}=1$ for $a=e$.

### 1.4.5. Infinitely Small and Infinitely Large Functions

**1. Infinitely small functions.**

**Definition 1.** *The function $f(x)$ is an infinitely small function (an infinitesimal) at the point $x=x_0$ (or as $x\to x_0$) if* $\lim\limits_{x\to x_0} f(x)=0$.

We can similarly define infinitely small functions as $x \to \infty$, $x \to +\infty$, $x \to -\infty$, $x \to x_0-$, $x \to x_0+$.

We can give an equivalent definition of an infinitely small function "in the ε-δ language": *the function $f(x)$ is infinitely small at the point $x = x_0$ if for any $\varepsilon > 0$ there is $\delta > 0$ such that the inequality $|f(x)| < \varepsilon$ holds true for all $x \in X$, $x \neq x_0$, which satisfy the inequality $|x - x_0| < \delta$* or, using the logical symbols,

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall x \in X,\ x \neq x_0,\ |x - x_0| < \delta):\ |f(x)| < \varepsilon,$$

and "in the language of sequences": *the function $f(x)$ is infinitely small at the point $x = x_0$ if for any, converging to $x_0$, sequence $\{x_n\}$ of the values of the argument $x$, which differ from $x_0$, the corresponding sequence $\{f(x_n)\}$ is infinitely small.*

Consider the following theorem.

**Theorem 4.5.** *For the equality* $\lim\limits_{x \to x_0} f(x) = A$ *to be satisfied, it is necessary and sufficient that the function*

$$\alpha(x) = f(x) - A$$

*be infinitely small as $x \to x_0$.*

*Proof.* **Necessity.** Let $\lim\limits_{x \to x_0} f(x) = A$. We shall consider the difference $f(x) - A = \alpha(x)$ and show that $\alpha(x)$ is an infinitesimal as $x \to x_0$. Indeed, as $x \to x_0$, the limits of every one of the functions $f(x)$ and $A$ are equal to $A$ and therefore, by virtue of Theorem 4.3,

$$\lim_{x \to x_0} \alpha(x) = \lim_{x \to x_0} [f(x) - A] = \lim_{x \to x_0} f(x) - \lim_{x \to x_0} A = A - A = 0.$$

**Sufficiency.** Let $f(x) - A = \alpha(x)$, where $\alpha(x)$ is an infinitesimal as $x \to x_0$. We shall prove that $\lim\limits_{x \to x_0} f(x) = A$. Since $f(x) = A + \alpha(x)$, we have

$$\lim_{x \to x_0} f(x) = \lim_{x \to x_0} [A + \alpha(x)] = \lim_{x \to x_0} A + \lim_{x \to x_0} \alpha(x) = A + 0 = A. \ \blacksquare$$

From Theorem 4.5 we get a special representation for the function which has a limit equal to $A$ at the point $x = x_0$:

$$f(x) = A + \alpha(x), \text{ where } \lim_{x \to x_0} \alpha(x) = 0.$$

In this case we usually say that in the neighbourhood of the point $x_0$ the function $f(x)$ differs from $A$ by an infinitely small function.

Infinitely small functions possess the same properties as infinitely small sequences.

The following theorem holds true.

**Theorem 4.6.** *As $x \to x_0$, the algebraic sum and the algebraic product of a finite number of infinitely small functions as well as the product of an infinitely small function by a bounded function are infinitely small functions.*

This theorem follows directly from the first definition of the limit of a function and Theorems 2.2-2.4.

Everything we have said about infinitesimal functions as $x \to x_0$ is valid for infinitesimal functions as $x \to \infty$, $x \to +\infty$, $x \to -\infty$, $x \to x_0-$, $x \to x_0+$.

**2. Infinitely large functions.**

**Definition 2.** *The function $f(x)$ is said to be an infinitely large function at the point $x = x_0$ (or as $x \to x_0$) if for any $\varepsilon > 0$ there is $\delta > 0$ such that the inequality $|f(x)| > \varepsilon$ holds true for all $x \in X$, $x \neq x_0$, which satisfy the inequality $|x - x_0| < \delta$.*

In this case we write $\lim\limits_{x \to x_0} f(x) = \infty$ and say that the function tends to infinity as $x \to x_0$, or that it has a limit infinity at the point $x = x_0$.

Now if the inequality $f(x) > \varepsilon$ ($f(x) < -\varepsilon$) is satisfied, then we write $\lim\limits_{x \to x_0} f(x) = +\infty$ $\left(\lim\limits_{x \to x_0} f(x) = -\infty\right)$ and say that the function has a limit infinity equal to $+\infty$ ($-\infty$) at the point $x_0$.

Using the logical symbols, we can write Definition 2 as

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall x \in X,\ x \neq x_0,\ |x - x_0| < \delta):\ |f(x)| > \varepsilon.$$

We define infinite one-sided limits by analogy with finite one-sided limits:

$$\lim_{x \to x_0+} f(x) = +\infty, \quad \lim_{x \to x_0+} f(x) = -\infty,$$

$$\lim_{x \to x_0-} f(x) = +\infty, \quad \lim_{x \to x_0-} f(x) = -\infty.$$

"In the $\varepsilon$-$\delta$ language" this definition is written as follows: $\lim\limits_{x \to x_0+} f(x) = +\infty$ *if for any $\varepsilon > 0$ there is $\delta > 0$ such that the inequality $f(x) > \varepsilon$ holds true for all $x \in X$ which satisfy the inequalities $x_0 < x < x_0 + \delta$.* The symbolic notation is

$$(\forall \varepsilon > 0)\ (\exists \delta > 0)(\ \forall x \in X,\ x_0 < x < x_0 + \delta):\ f(x) > \varepsilon,$$

and "in the language of sequences": $\lim\limits_{x \to x_0+} f(x) = +\infty$ *if for any, converging to $x_0$, sequence $\{x_n\}$ of the values of the argument $x$, whose elements $x_n$ are greater than $x_0$, the corresponding sequence $\{f(x_n)\}$ of the values of the function is an infinitely large sequence of a positive sign.*

We invite the reader to give exact definitions of other similar limits.

We can give similar definitions of infinitely large functions as $x \to \infty$, $x \to +\infty$, $x \to -\infty$. Thus, for instance, the function $f(x)$ is infinitely large as $x \to \infty$ if for any $\varepsilon > 0$ there is $\delta > 0$ such that the inequality $|f(x)| > \varepsilon$ holds true for all $x \in X$ which satisfy the inequality $|x| > \delta$. In this case we write $\lim\limits_{x \to \infty} f(x) = \infty$.

When $x \to \infty$, the symbolic notation of the definition of an infinitely large function is

$$(\forall \varepsilon > 0)\ (\exists \delta > 0)\ (\forall x \in X,\ |x| > \delta):\ |f(x)| > \varepsilon.$$

Now if the inequality $f(x) > \varepsilon$ ($f(x) < -\varepsilon$) is satisfied, then we write $\lim\limits_{x \to \infty} f(x) = +\infty$ ($\lim\limits_{x \to \infty} f(x) = -\infty$).

We invite the reader to formulate the definition of an infinitely large function as $x \to +\infty$, $x \to -\infty$.

We shall show in conclusion that there is the same relationship between infinitely small and infinitely large functions as between the respective sequences, i.e. the inverse of an infinitely small function is an infinitely large function and vice versa.

Indeed, let $\lim\limits_{x \to x_0} f(x) = 0$ and $f(x) \neq 0$ as $x \neq x_0$. We shall prove that $\lim\limits_{x \to x_0} \frac{1}{f(x)} = \infty$. We take an arbitrary $\varepsilon > 0$. Since $f(x)$ is an infinitesimal function at the point $x_0$, it follows that for the number $1/\varepsilon$ there is $\delta > 0$ such that the inequality $|f(x)| < 1/\varepsilon$ holds true for all $x \in X$ which satisfy the inequalities $0 < |x - x_0| < \delta$. But then the inequality $|1/f(x)| > \varepsilon$ is satisfied for the same $x$, i.e. $1/f(x)$ is an infinitely large function at the point $x = x_0$, and this is what we wished to prove.

We invite the reader to formulate the converse definition.

### 1.4.6. Comparing Infinitely Small and Infinitely Large Functions

We have shown that the sum, the difference and the product of infinitesimal functions are infinitesimal functions. We cannot, in general, say this about the quotient: the division of one infinitesimal by another may lead to different results. Thus, for instance, if $\alpha(x) = x$ and $\beta(x) = 2x$, then

$$\lim_{x \to 0} \frac{\alpha(x)}{\beta(x)} = \lim_{x \to 0} \frac{x}{2x} = \frac{1}{2}.$$

Now if $\alpha(x) = x$ and $\beta(x) = x^2$, then

$$\lim_{x \to 0} \frac{\alpha(x)}{\beta(x)} = \lim_{x \to 0} \frac{1}{x} = \infty, \quad \lim_{x \to 0} \frac{\beta(x)}{\alpha(x)} = \lim_{x \to 0} x = 0.$$

Let us consider the rules of comparing infinitesimals.

Assume that as $x \to x_0$, the functions $\alpha(x)$ and $\beta(x)$ are infinitesimal. Then

(1) if $\lim\limits_{x \to x_0} \frac{\alpha(x)}{\beta(x)} = 0$, then $\alpha(x)$ is an infinitesimal of a higher order than $\beta(x)$ (we also say that $\alpha(x)$ has a higher order of smallness than $\beta(x)$ as $x \to x_0$);

(2) if $\lim\limits_{x \to x_0} \frac{\alpha(x)}{\beta(x)} = A \neq 0$ ($A$ is a number), then $\alpha(x)$ and $\beta(x)$ are infinitesimals of the same order;

(3) if $\lim\limits_{x \to x_0} \frac{\alpha(x)}{\beta(x)} = 1$, then $\alpha(x)$ and $\beta(x)$ are equivalent infinitesimals. The equivalence is designated as $\alpha(x) \sim \beta(x)$.

In some cases it is insufficient to know that one of the two infinitesimals is of a higher order of smallness than the other. It is also necessary to find how high the order of smallness is. Therefore the following rule is introduced:

(4) if $\lim\limits_{x \to x_0} \frac{\alpha(x)}{\beta^n(x)} = A \neq 0$, then $\alpha(x)$ is an infinitesimal of order $n$ relative to $\beta(x)$.

There are similar rules of comparing infinitesimals as $x \to \infty$, $x \to -\infty$, $x \to +\infty$, and also as $x \to x_0$ on the right and on the left.

Here are some examples.

1. As $x \to 0$, the functions $\sin x$ and $x$ are equivalent infinitesimals since $\lim\limits_{x \to 0} \frac{\sin x}{x} = 1$.

2. As $x \to 0$, the functions $\sin 3x$ and $\sin x$ are infinitesimals of the same order since

$$\lim_{x \to 0} \frac{\sin 3x}{\sin x} = \lim_{x \to 0} \frac{(3 \sin 3x)/(3x)}{(\sin x)/x} = 3 \lim_{x \to 0} \frac{\sin 3x}{3x} \lim_{x \to 0} \frac{x}{\sin x} = 3.$$

3. As $x \to 0$, the function $\alpha(x) = 1 - \cos x$ is an infinitesimal of the second order of smallness relative to the infinitesimal $x$ since

$$\lim_{x \to 0} \frac{1 - \cos x}{x^2} = \lim_{x \to 0} \frac{2 \sin^2 (x/2)}{x^2} = \frac{1}{2} \lim_{x \to 0} \left( \frac{\sin (x/2)}{x/2} \right)^2 = \frac{1}{2}.$$

When comparing infinitesimals, we often use the symbol $o$ small. If, at the point $x_0$, the function $\alpha(x)$ is an infinitesimal of the order of smallness higher than that of an infinitesimal $\beta(x)$ at the same point, the conventional notation is

$$\alpha(x) = o(\beta(x)).$$

If the functions $\alpha(x)$ and $\beta(x)$ are infinitesimals at the point $x_0$, then the function $\alpha(x)\beta(x)$ has a higher order of smallness than

each of the factors. Indeed,

$$\lim_{x\to x_0} \frac{\alpha(x)\,\beta(x)}{\beta(x)} = \lim_{x\to x_0} \alpha(x) = 0$$

and therefore $\alpha(x)\,\beta(x) = o(\beta(x))$, $\alpha(x)\,\beta(x) = o(\alpha(x))$.

If $\alpha(x) \sim \alpha_1(x)$ and $\beta(x) \sim \beta_1(x)$ as $x \to x_0$ and there is $\lim\limits_{x\to x_0} \frac{\alpha(x)}{\beta(x)}$, then there also exists $\lim\limits_{x\to x_0} \frac{\alpha_1(x)}{\beta_1(x)}$, and $\lim\limits_{x\to x_0} \frac{\alpha_1(x)}{\beta_1(x)} = \lim\limits_{x\to x_0} \frac{\alpha(x)}{\beta(x)}$. Indeed, we have

$$\lim_{x\to x_0} \frac{\alpha_1(x)}{\beta_1(x)} = \lim_{x\to x_0} \left[ \frac{\alpha_1(x)}{\alpha(x)} \frac{\alpha(x)}{\beta(x)} \frac{\beta(x)}{\beta_1(x)} \right]$$

$$= \lim_{x\to x_0} \frac{\alpha_1(x)}{\alpha(x)} \lim_{x\to x_0} \frac{\alpha(x)}{\beta(x)} \lim_{x\to x_0} \frac{\beta(x)}{\beta_1(x)} = 1 \cdot \lim_{x\to x_0} \frac{\alpha(x)}{\beta(x)} \cdot 1 = \lim_{x\to x_0} \frac{\alpha(x)}{\beta(x)}.$$

In many cases the assertion we have proved simplifies the calculation of limits.

**Example.** Find $\lim\limits_{x\to 0} \frac{\sin 5x}{x+x^3}$.

*Solution.* Since $\sin 5x \sim 5x$, $x + x^3 \sim x$ as $x \to 0$, we have

$$\lim_{x\to 0} \frac{\sin 5x}{x+x^3} = \lim_{x\to 0} \frac{5x}{x} = 5.$$

Similar rules of comparison hold true for infinitely large functions. Let us consider some examples.

1. As $x \to 0$, the functions $\alpha(x) = (1+x)/x$ and $\beta(x) = 1/x$ are equivalent infinitely large functions since

$$\lim_{x\to 0} \frac{\alpha(x)}{\beta(x)} = \lim_{x\to 0} (1+x) = 1.$$

In this case we also say that $\alpha(x)$ and $\beta(x)$ have the same order of increase as $x \to 0$.

2. As $x \to \infty$, the function $\alpha(x) = x^2 + 4$ is an infinitely large function of the order lower than that of $\beta(x) = x^3 - 2$ (has a lower order of increase) since

$$\lim_{x\to\infty} \frac{\alpha(x)}{\beta(x)} = \lim_{x\to\infty} \frac{x^2+4}{x^3-2} = \lim_{x\to\infty} \frac{1+4/x^2}{x-2/x^2} = \lim_{x\to\infty} \frac{1}{x} = 0.$$

3. As $x \to \infty$, infinitely large functions $\alpha(x) = 2x^2 + 1$ and $\beta(x) = x^2 - 1$ have the same order of increase since

$$\lim_{x\to\infty} \frac{2x^2+1}{x^2-1} = \lim_{x\to\infty} \frac{2+1/x^2}{1-1/x^2} = 2.$$

4. As $x \to \infty$, the function $\alpha(x) = x^4 + x + 1$ is an infinitely large function of the second order relative to the infinitely large function $\beta(x) = x^2 + 1$ since

$$\lim_{x\to\infty} \frac{x^4+x+1}{(x^2+1)^2} = \lim_{x\to\infty} \frac{x^4+x+1}{x^4+2x^2+1} = \lim_{x\to\infty} \frac{1+1/x^3+1/x^4}{1+2/x^2+1/x^4} = 1.$$

### 1.4.7. Continuity of a Function

The concept of the continuity of a function is one of the basic concepts in analysis.

**1. The definition of the continuity of a function.** Let the function $f(x)$ be defined in some neighbourhood of the point $x_0$.

**Definition 1.** *The function $f(x)$ is continuous at the point $x_0$ if the limit of the function and its value at that point are equal, i.e.*

$$\lim_{x\to x_0} f(x) = f(x_0)\,^{*}. \tag{1}$$

Since $\lim\limits_{x\to x_0} x = x_0$, relation (1) can be written in the form

$$\lim_{x\to x_0} f(x) = f\left(\lim_{x\to x_0} x\right),$$

i.e. we can interchange the sign of the function and that of the limit for a continuous function.

Here is an equivalent definition of the continuity of a function "in the language of sequences": *the function $f(x)$ is continuous at the point $x_0$ if for any sequence of the values of the argument $x$: $x_1, x_2, x_3, \ldots, x_n, \ldots$, which converges to $x_0$, the sequence of the corresponding values of the function $f(x_1), f(x_2), f(x_3), \ldots, f(x_n), \ldots$ converges to $f(x_0)$.*

Here is the definition of the continuity of a function "in the $\varepsilon$-$\delta$ language".

**Definition 2.** *The function $f(x)$ is continuous at the point $x_0$ if for any $\varepsilon > 0$ there is $\delta > 0$ such that the inequality $|f(x) - f(x_0)| < \varepsilon$ holds true for all $x$ which satisfy the inequality $|x - x_0| < \delta$.*

The equivalence of these definitions is obvious.

Let us write Definition 2 using logical symbols:

$$(\forall \varepsilon > 0)\,(\exists \delta > 0)\,(\forall x \in X,\ |x - x_0| < \delta):\ |f(x) - f(x_0)| < \varepsilon.$$

If $\lim\limits_{x\to x_0+} f(x) = f(x_0)$ $\left(\lim\limits_{x\to x_0-} f(x) = f(x_0)\right)$, then the function $f(x)$ is said to be *continuous at the point $x_0$ from the right (from the left)*. If the function $f(x)$ is continuous at the point $x_0$ both from the left

* It follows from the definition that if a function is continuous at the point $x_0$, then it is defined at this point, i.e. there is $f(x_0)$. Note that this was not required when we defined the limit of a function at the point $x_0$.

and from the right, then it is continuous at this point. Indeed, in this case, by virtue of Theorem 4.2, the limit of the function at the point $x_0$ is equal to its value at this point.

Here is one more definition of the continuity of a function which is in fact the paraphrase of the first definition. In relation (1) we transfer $f(x_0)$ to the left-hand side and put $f(x_0)$ under the sign of the limit. The conditions $x \to x_0$ and $(x - x_0) \to 0$ being equivalent, we obtain

$$\lim_{(x-x_0)\to 0} [f(x) - f(x_0)] = 0. \tag{2}$$

The difference $x - x_0$ is the *increment of the argument $x$ at the point $x_0$* and is designated, as a rule, as $\Delta x$, and the difference $f(x) - f(x_0)$ is the *increment of the function at the point $x_0$* caused by the increment of the argument $\Delta x$ and is designated as $\Delta y$. Thus

$$\Delta x = x - x_0, \quad \Delta y = f(x_0 + \Delta x) - f(x_0).$$

Note that when the point $x_0$ is fixed, $\Delta y$ is the function of the argument $\Delta x$. The geometric meaning of increments is clear from Fig. 50. In the new designations, relation (2) assumes the form

$$\lim_{\Delta x \to 0} \Delta y = 0. \tag{3}$$

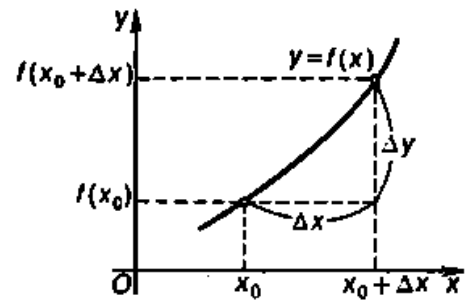


Fig. 50

Relation (3) is one more definition of the continuity of a function, which can be formulated as follows.

**Definition 3.** *The function $f(x)$ is continuous at the point $x_0$ if its increment at this point is an infinitesimal function as $\Delta x \to 0$.*

For practical applications the last definition is often more convenient and we shall also use it.

**2. Arithmetic operations involving continuous functions.**

**Theorem 4.7.** *Let the functions $f(x)$ and $g(x)$ be continuous at the point $x_0$. Then the functions $f(x) \pm g(x)$, $f(x)\,g(x)$, $\frac{f(x)}{g(x)}$ are also continuous at this point (the latter for $g(x_0) \neq 0$).*

*Proof.* Since the functions $f(x)$ and $g(x)$ continuous at the point $x_0$ have limits equal to $f(x_0)$ and $g(x_0)$ at this point, it follows, according to Theorem 4.3, that the limits of the functions $f(x) \pm g(x)$, $f(x)\,g(x)$, $\frac{f(x)}{g(x)}$ exist and are equal to $f(x_0) \pm g(x_0)$, $f(x_0)\,g(x_0)$, $\frac{f(x_0)}{g(x_0)}$ respectively. But these quantities are equal to the values of the corresponding functions at the point $x_0$. Consequently, according to Definition 1, the functions $f(x) \pm g(x)$, $f(x)\,g(x)$, $\frac{f(x)}{g(x)}$ are continuous at the point $x_0$. ■

### 1.4.8. Continuity of Some Elementary Functions

One of the most important properties of elementary functions is their continuity at every point in whose neighbourhood they are defined. We shall consider some functions to verify this fact using the definition of the continuity of a function at a point and Theorem 4.7.

**1. Continuity of rational functions.** The constant function $f(x) = C$ can serve as the simplest example of a function continuous at any point $x_0$ of the number line. Indeed, in this case $\lim\limits_{x \to x_0} f(x) = C = f(x_0)$ (see Example 1 in 1.4.2), i.e. a constant function is continuous at every point of the number line.

The function $f(x) = x$ is also continuous at every point $x_0$ of the number line since $\lim\limits_{x \to x_0} x = x_0 = f(x_0)$ (see Example 2 in 1.4.2), i.e. the limit of a function at the point $x_0$ is equal to its value at this point. From what we have said above and from Theorem 4.7 it follows that at any point $x_0$ the functions $x^2 = x \cdot x$, $x^3 = x^2 \cdot x$, $x^4 = x^3 \cdot x$, . . ., $x^n = x^{n-1} \cdot x$ ($n$ is a natural number) are continuous. As is known, the function $f(x) = x^n$ is a power function and a function of the form

$$P(x) = C_0 x^n + C_1 x^{n-1} + C_2 x^{n-2} + \ldots + C_{n-1} x + C_n,$$

where $n \geqslant 0$ is an integer and $C_0, C_1, C_2, \ldots, C_n$ are any numbers, is an algebraic polynomial.

Each of the terms

$$C_0 x^n, \ C_1 x^{n-1}, \ C_2 x^{n-2}, \ \ldots, \ C_n$$

is the product of two continuous functions (a constant function and a power function). According to Theorem 4.7, it is continuous at any point $x$. Thus the polynomial $P(x)$ is the sum of the functions which are continuous at any point $x$ and, consequently, is continuous at any point $x$.

A fractional rational function, i.e. a function of the form

$$R(x) = \frac{P(x)}{Q(x)},$$

where $P(x)$ and $Q(x)$ are algebraic polynomials, is continuous at all points $x$ at which its denominator is nonzero (i.e. at all points except for the roots of the denominator), as a quotient of continuous functions. For example, the function $R(x) = (3x^2 + 7x - 1)/(x^2 - 1)$ is continuous at all points $x$ different from $+1$ and $-1$.

**2. Continuity of trigonometric functions.** Let us consider trigonometric functions $\sin x$, $\cos x$, $\tan x$, $\cot x$, $\sec x$, $\operatorname{cosec} x$. We shall show that the function $\sin x$ is continuous at any point $x$. We make

use of Definition 3 of the continuity of a function. Assigning an increment $\Delta x$ to the argument $x$, we get the increment of the function

$$\Delta y = \sin(x + \Delta x) - \sin x,$$

or

$$\Delta y = 2\cos\left(x + \frac{\Delta x}{2}\right)\sin\frac{\Delta x}{2}.$$

Passing to the limit on the left-hand and right-hand sides of the relation as $\Delta x \to 0$, we get

$$\lim_{\Delta x \to 0} \Delta y = 2 \lim_{\Delta x \to 0}\left[\cos\left(x + \frac{\Delta x}{2}\right)\sin\frac{\Delta x}{2}\right] = 0$$

since $|\cos(x + \Delta x/2)| \leqslant 1$,

$$\lim_{\Delta x \to 0} \sin\frac{\Delta x}{2} = \lim_{\Delta x \to 0}\frac{\sin(\Delta x/2)}{\Delta x/2}\lim_{\Delta x \to 0}\frac{\Delta x}{2} = \frac{1}{2}\cdot 1\cdot \lim_{\Delta x \to 0}\Delta x = 0\,*,$$

and the product of a bounded function by an infinitesimal is an infinitesimal. Thus the function $\sin x$ is continuous at any point $x$.

The continuity of the function $\cos x$ at any point $x$ can be proved by analogy.

According to Theorem 4.7, the continuity of the functions $\sin x$ and $\cos x$ implies the continuity of the functions $\tan x = \sin x/\cos x$ and $\sec x = 1/\cos x$ at all points where $\cos x \neq 0$, i.e. at all points except for $x = \pi/2 + n\pi$, and the continuity of the functions $\cot x = \cos x/\sin x$ and $\operatorname{cosec} x = 1/\sin x$ at all points except for $x = n\pi$ ($n = 0, \pm 1, \pm 2, \ldots$).

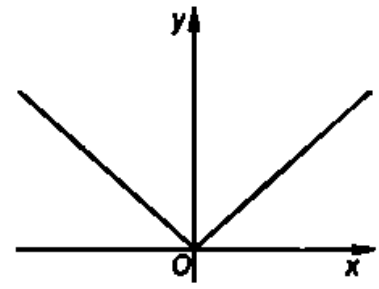


Fig. 51

**3. Continuity of the function $f(x) = |x|$.** The function $f(x) = |x|$, whose graph is shown in Fig. 51, is defined and continuous at all points of the number line. Indeed, it is continuous at the points of the ray $(0, +\infty)$ since $f(x) = x$ for $x > 0$ (see Item 1). The function $f(x)$ is also continuous at the points of the ray $(-\infty, 0)$ since $f(x) = -x$ for $x < 0$ and we can represent it as the product of two continuous functions $(-1)$ and $x$ and apply Theorem 4.7 on the continuity of the product. To establish the continuity of the function $|x|$ at the point $x = 0$, we shall calculate the one-sided limits of the function at this points:

$$\lim_{x \to 0-}|x| = \lim_{x \to 0-}(-x) = -\lim_{x \to 0-} x = 0, \quad \lim_{x \to 0+}|x| = \lim_{x \to 0+} x = 0.$$

* We have used here the first remarkable limit which results from the change of variable $t = \Delta x/2$: $\lim\limits_{\Delta x \to 0}\dfrac{\sin \Delta x/2}{\Delta x/2} = \lim\limits_{t \to 0}\dfrac{\sin t}{t} = 1$ (it is evident that $t = \Delta x/2 \to 0$ as $\Delta x \to 0$).

Thus the left-hand and right-hand limits of the function at the point $x = 0$ coincide and are equal to the value of the function at this point. Hence the function $|x|$ is continuous at the point $x = 0$ and, consequently, continuous at all points of the number line.

Thus the functions we have considered are continuous at every point in whose neighbourhood they are defined. On the basis of Theorem 4.7 on the continuity of the sum, the difference, the product and the quotient, we can state that functions obtained from them by means of a finite number of arithmetic operations are also continuous functions at every point in whose neighbourhood they are defined.

We say that the function $f(x)$ is continuous in the open interval $(a, b)$ if it is continuous at every point of the interval, that it is continuous in the closed interval $[a, b]$ if it is continuous in the open interval $(a, b)$ and continuous from the right at the point $a$ and from the left at the point $b$, i.e.

$$\lim_{x\to a+} f(x) = f(a), \quad \lim_{x\to b-} f(x) = f(b).$$

### 1.4.9. Classification of Points of Discontinuity of Functions

**1. Definition and classification of points of discontinuity of functions.**

**Definition.** *The point $x_0$ is a point of discontinuity of the function $f(x)$ if $f(x)$ is not continuous at the point $x_0$.*

The discontinuities of functions are classified as follows.

**Discontinuity of the first kind.** The point $x_0$ is a *point of discontinuity of the first kind* of the function $f(x)$ if at this point the function $f(x)$ has finite unequal right-hand and left-hand limits:

$$\lim_{x\to x_0+} f(x) \neq \lim_{x\to x_0-} f(x).$$

**Example.** For the function $f(x) = \operatorname{sgn} x$ the point $x = 0$ is a point of discontinuity of the first kind (see Fig. 46) since $\lim\limits_{x\to 0+} \operatorname{sgn} x = 1$ and $\lim\limits_{x\to 0-} \operatorname{sgn} x = -1$.

**Discontinuity of the second kind.** The point $x_0$ is a *point of discontinuity of the second kind* of the function $f(x)$ if at this point the function $f(x)$ does not have at least one of the one-sided limits or if at least one of the one-sided limits is infinite.

**Example.** For the function $f(x) = 1/x$ the point $x = 0$ is a point of discontinuity of the second kind (see Fig. 48) since $\lim\limits_{x\to 0+} (1/x) = +\infty$ and $\lim\limits_{x\to 0-} (1/x) = -\infty$.

**2. Piecewise continuous functions.** The function $f(x)$ is *piecewise continuous on the closed interval* $[a, b]$ if it is continuous at all interior points of $[a, b]$, except maybe for a finite number of points at which it has a discontinuity of the first kind and, in addition, has one-sided limits at the points $a$ and $b$.

A function is *piecewise continuous on the number line* if it is piecewise continuous on any closed interval.

**Example.** The function $f(x) = [x]$ is piecewise continuous both on any closed interval and on the entire number line. Recall that the symbol $[x]$ denotes the integral part of the number $x$. The graph of the function $f(x) = [x]$ is shown in Fig. 52, the function $[x]$ is continuous from the right and discontinuous from the left at the points $x = n$ $(n = 0, \pm 1, \pm 2, \ldots)$. At all other points it is continuous from the right and from the left.

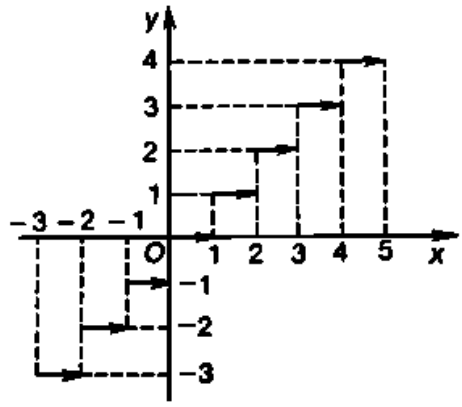


Fig. 52

## 1.4.10. Principal Properties of Continuous Functions

**1. Theorem on the constancy of sign of a continuous function.**

**Theorem 4.8.** *Let the function $f(x)$ be continuous at the point $x_0$ and $f(x_0) \neq 0$. Then there is $\delta > 0$ such that for all $x \in (x_0 - \delta, x_0 + \delta)$ the function $f(x)$ has the same sign as $f(x_0)$.*

*Proof.* Let $f(x_0) > 0$ (Fig. 53). Then, by virtue of the second definition of the continuity of a function, for any $\varepsilon > 0$ there is

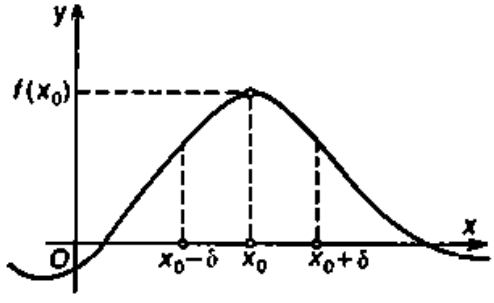


Fig. 53

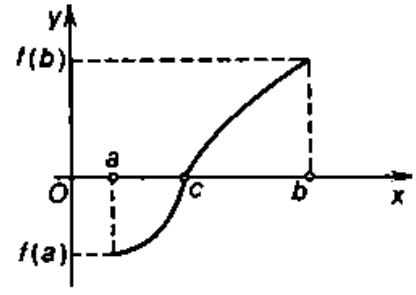


Fig. 54

$\delta > 0$ such that the inequality $|f(x) - f(x_0)| < \varepsilon$ holds true for all $x$ which satisfy the condition $|x - x_0| < \delta$, or, what is the same, the inequalities

$$f(x_0) - \varepsilon < f(x) < f(x_0) + \varepsilon \tag{1}$$

hold true for all $x \in (x_0 - \delta, x_0 + \delta)$. We take $\varepsilon = f(x_0)$. Then, from the left-hand inequality (1) we find that $f(x) > 0$ for all $x \in (x_0 - \delta, x_0 + \delta)$, and this is what we wished to prove.

Now if $f(x_0) < 0$, then we consider a function $-f(x)$. Since $-f(x_0) > 0$, it follows, from what we have proved, that there is a $\delta$-neighbourhood of the point $x_0$ at which $-f(x) > 0$ and, consequently, $f(x) < 0$. ■

**2. A continuous function passing through any intermediate value.** Let us consider the theorem on a continuous function passing through a zero value when the signs are changed.

**Theorem 4.9 (Bolzano*-Cauchy first theorem).** *Let the function $f(x)$ be continuous on the closed interval $[a, b]$ and have values of opposite signs at the endpoints of the interval. Then there is a point $c \in (a, b)$ at which $f(c) = 0$.*

*Proof.* Assume, for definiteness, that $f(a) < 0$ and $f(b) > 0$ (Fig. 54). We bisect the closed interval $[a, b]$. If the value of the function is zero at the midpoint of the interval $[a, b]$, then the theorem is proved. Otherwise we choose one of the two intervals obtained at whose endpoints the function has values of opposite signs and designate it as $[a_1, b_1]$. We bisect the interval $[a_1, b_1]$. If the value of the function is zero at the midpoint of the interval $[a_1, b_1]$, then the theorem is proved. Otherwise we choose one of the two intervals obtained at whose endpoints the function $f(x)$ has values of opposite signs and designate it as $[a_2, b_2]$. If we continue this process indefinitely, then either at some $k$th step the value of the function will prove to be zero at the midpoint of the interval $[a_k, b_k]$ and then the theorem will be proved, or we shall get a sequence

$$[a, b] \supset [a_1, b_1] \supset [a_2, b_2] \supset \ldots \supset [a_n, b_n] \supset \ldots$$

of nested intervals. In this case $b_n - a_n = (b - a)/2^n \to 0$ as $n \to \infty$ and the function has values of opposite signs at the endpoints of each interval $[a_n, b_n]$.

According to Theorem 2.13 on nested intervals, there is a point $c$ which belongs to all the intervals. We shall prove that $f(c) = 0$. Indeed, if we assume that $f(c) > 0$, then, according to Theorem 4.8 on the constancy of sign of a continuous function, there is a neighbourhood of the point $c$ in which $f(x) > 0$. For a sufficiently large $n$ the interval $[a_n, b_n]$ falls in this neighbourhood and, consequently, the inequality $f(x) > 0$ is satisfied on the interval $[a_n, b_n]$. But this contradicts the fact that at the endpoints of the interval $[a_n, b_n]$ the function has values of opposite signs. We can prove by analogy that $f(c)$ cannot be smaller than zero. It remains for us to accept that $f(c) = 0$. In this case it is evident that the point $c \in (a, b)$. ■

* Bolzano, Bernhard (1781-1848), a Czechoslovakian mathematician and logician.

The theorem we have proved has a simple geometric meaning: when a continuous curve passes from one half-plane, whose boundary is the abscissa axis, to the other half-plane, it cuts that axis.

Here is the theorem on a continuous function passing through any intermediate value.

**Theorem 4.10 (Bolzano-Cauchy second theorem).** *Assume that the function $f(x)$ is continuous on the closed interval $[a, b]$ and $f(a) = A$ and $f(b) = B$. Assume, furthermore, that $C$ is any number included between $A$ and $B$. Then there is a point $c$ on the interval $[a, b]$ such that $f(c) = C$.*

In other words, when a continuous function passes from one value to another, it also assumes all intermediate values.

*Proof.* Assume, for definiteness, that $A < B$ and $A < C < B$ (Fig. 55). We consider an auxiliary function

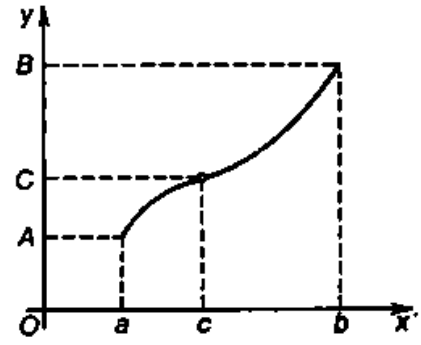


Fig. 55

$$\varphi(x) = f(x) - C.$$

This function is continuous on the closed interval $[a, b]$ (being the difference of continuous functions) and assumes values of opposite signs at the endpoints of the interval:

$$\varphi(a) = f(a) - C = A - C < 0,$$
$$\varphi(b) = f(b) - C = B - C > 0.$$

According to Theorem 4.9, there is a point $c \in (a, b)$ such that $\varphi(c) = f(c) - C = 0$. Hence $f(c) = C$. ■

**Corollary.** *If the function $f(x)$ is defined and continuous on some interval $X$, then the set of its values $Y$ is also an interval.*

Before proving this corollary, we shall introduce the concepts of the least upper bound and the greatest lower bound of a function. Assume that the function $y = f(x)$ is defined on the set $X$ and that $Y$ is its range. If the set $Y$ is bounded from above (from below), then it has the least upper bound (the greatest lower bound). The least upper bound (the greatest lower bound) of the set $Y$ is known as the *least upper bound*, or *supremum* (the *greatest lower bound*, or *infimum*) *of the function* $y = f(x)$ on the set $X$ and is designated as $\sup\limits_X f(x)$ ($\inf\limits_X f(x)$). In other words, we can formulate the definitions of the least upper bound (the greatest lower bound) of the function $y = f(x)$ on the set $X$ as follows: *the number $M$ ($m$) is the least upper bound (the greatest lower bound) of the function $y = f(x)$ on the set $X$ if the following two conditions are satisfied:*

(1) $f(x) \leqslant M$ ($f(x) \geqslant m$) *for any* $x \in X$;

(2) *for any number* $M' < M$ $(m' > m)$ *there is a point* $x' \in X$ *such that* $f(x') > M'$ $(f(x') < m')$.

The first condition shows that the number $M$ $(m)$ is one of the upper (lower) bounds of the function $y = f(x)$ on the set $X$ and the second condition shows that $M$ $(m)$ is the least (greatest) upper (lower) bound of the function.

If the set $Y$ is not bounded from above (from below), then we write $\sup\limits_X f(x) = +\infty$ $(\inf\limits_X f(x) = -\infty)$. In this case for any number $A$ there is a point $x' \in X$ such that $f(x') > A$ $(f(x') < A)$.

We shall now prove the corollary of Theorem 4.10.

*Proof.* Let $m = \inf\limits_X f(x)$ and $M = \sup\limits_X f(x)$. We take any $y$ from $Y$ which is not equal to $m$ or $M$ and choose two values $y_1$ and $y_2$ of the function $f(x)$ such that the inequalities

$$m \leqslant y_1 < y < y_2 \leqslant M$$

are satisfied (if $M = +\infty$ $(m = -\infty)$, then $y_2 < M$ $(m < y_1)$). The existence of these values of the function $f(x)$ follows from the definitions of the supremum and infimum. Then, in accordance with Theorem 4.10 on the intermediate values of a continuous function, there is a point $x$ such that $f(x) = y$. Consequently, the set $Y$ is an interval (finite or infinite) with endpoints $m$ and $M$ which, depending on a specific case, may belong to it or not belong to it. ■

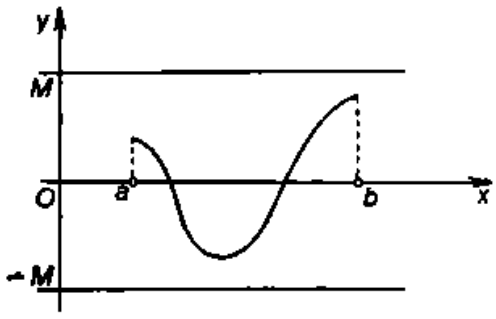


Fig. 56

**3. Theorem on the boundedness of a continuous function on a closed interval.** Recall that the function $f(x)$ is said to be bounded on the closed interval $[a, b]$ if there is a number $M > 0$ such that the inequality $|f(x)| \leqslant M$ or $-M \leqslant f(x) \leqslant M$ is satisfied for all $x \in [a, b]$, i.e. the graph of $f(x)$ does not leave the strip bounded by the straight lines $y = M$ and $y = -M$ (Fig. 56).

**Theorem 4.11 (Weierstrass's* first theorem).** *If the function* $f(x)$ *is defined and continuous on the closed interval* $[a, b]$, *then it is bounded on this interval.*

We shall first prove the following lemma.

**Lemma.** *The function* $f(x)$, *which is continuous at the point* $x_0$, *is bounded in its certain neighbourhood.*

*Proof.* Let $\varepsilon = 1$. Then, in accordance with the second definition of the continuity of a function at a point, for this $\varepsilon$ there is $\delta > 0$ such that the inequality $|f(x) - f(x_0)| < 1$ holds true for all

* Weierstrass, Karl Theodor Wilhelm (1815-1897), a German mathematician.

$x \in (x_0 - \delta,\ x_0 + \delta)$. Using this inequality, we get $|f(x)| = |(f(x) - f(x_0)) + f(x_0)| \leqslant |f(x) - f(x_0)| + |f(x_0)| < 1 + |f(x_0)|$, i.e. $|f(x)| < M$, where $M = 1 + |f(x_0)|$. Hence we infer that the function $f(x)$ is bounded in the $\delta$-neighbourhood of the point $x_0$. ■

*Proof of the theorem.* We assume the contrary, i.e. suppose that the function $f(x)$ is unbounded on the closed interval $[a, b]$. We divide the interval $[a, b]$ in two, and then at least on one of the two intervals obtained the function $f(x)$ is unbounded (otherwise it would be bounded on $[a, b]$). We designate this interval as $[a_1, b_1]$. We divide $[a_1, b_1]$ in two and designate the interval on which the function $f(x)$ is unbounded as $[a_2, b_2]$, and so on. Continuing this process indefinitely, we get a sequence

$$[a, b] \supset [a_1, b_1] \supset [a_2, b_2] \supset \ldots \supset [a_n, b_n] \supset \ldots$$

of nested intervals on each of which $f(x)$ is unbounded, and $b_n - a_n = (b - a)/2^n \to 0$ as $n \to \infty$.

According to Theorem 2.13 on nested intervals, there is a point $c$ which belongs to all the intervals. By the hypothesis, the function $f(x)$ is defined and continuous at the point $c$ and, consequently, according to the lemma we have proved, it is bounded in some neighbourhood of the point $c$. For a sufficiently large $n$ the interval $[a_n, b_n]$ on which the function $f(x)$ is also bounded falls in this neighbourhood. But this contradicts the fact that $f(x)$ is unbounded on each of the nested intervals. The contradiction obtained proves the theorem. ■

**Remark.** The theorem does not hold true if the closed interval $[a, b]$ is replaced by the open interval $(a, b)$. Thus, for instance, the function $f(x) = 1/x$ is continuous on $(0, 1)$ but is unbounded since $\lim\limits_{x \to 0+} (1/x) = +\infty$. We cannot prove the theorem for an open interval when it is stated that the function is defined and continuous at the point $c$. For an open interval the point $c$ may coincide with its endpoint and then $f(x)$ will not be defined and continuous at the point $c$.

**4. Theorem stating that a function continuous on a closed interval attains its supremum and infimum.** In the case when the least upper bound and the greatest lower bound of a function are its values, we say that the function attains its supremum and infimum. However (see formula (1) in Chapter 1 and Theorem 1.1), not every set includes its supremum and infimum. The following example shows that the least upper bound and the greatest lower bound are not always attained.

Assume that the function $f(x) = x - [x]$, whose graph is shown in Fig. 57, is defined on the interval $[0, b]$, $b \geqslant 1$. The interval $[0, 1)$ is its range. The function is bounded both from above and from

below, has the least upper bound, equal to 1, and the greatest lower bound, equal to 0, on this interval. The function evidently assumes the value 0 but does not assume the value 1. Consequently, we can say that the function attains its greatest lower bound and does not attain its least upper bound.

Let us find the condition under which the function attains its supremum and infimum.

**Theorem 4.12 (Weierstrass's second theorem).** *If the function $f(x)$ is continuous on the closed interval $[a,b]$, then it attains its least upper bound and greatest lower bound on this interval, i.e. there are points $x_1, x_2 \in [a, b]$ such that* (Fig. 58)

$$f(x_1) = M = \sup_{[a,\,b]} f(x), \quad f(x_2) = m = \inf_{[a,\,b]} f(x).$$

*Proof.* Since the function $f(x)$ is continuous on the closed interval $[a, b]$, it follows, in accordance with Theorem 4.11, that it is

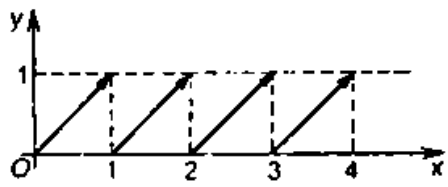



Fig. 57

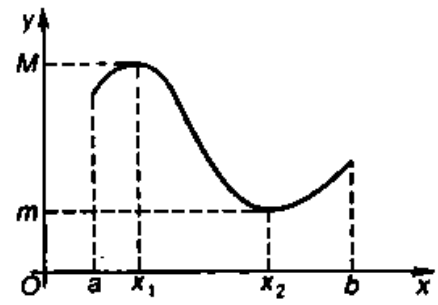



Fig. 58

bounded on this interval. Consequently, according to Theorem 1.1, the function $f(x)$ has the least upper $M$ and the greatest lower $m$ bound on the interval $[a, b]$.

We shall show that the function $f(x)$ attains $M$, i.e. there is a point $x_1 \in [a, b]$ such that $f(x_1) = M$. We shall use the rule of contraries. We assume that the function $f(x)$ does not assume the value equal to $M$ at any point of $[a, b]$. Then the inequality $f(x) < M$ is valid for all $x \in [a, b]$. We consider on $[a, b]$ an auxiliary, everywhere positive function

$$F(x) = \frac{1}{M - f(x)}.$$

According to Theorem 4.7, the function $F(x)$ is continuous as the quotient of two continuous functions. In this case, according to Theorem 4.11, the function $F(x)$ is bounded, i.e. there is a positive number $\mu$ such that

$$F(x) = \frac{1}{M - f(x)} \leqslant \mu, \text{ whence } f(x) \leqslant M - \frac{1}{\mu}$$

for all $x \in [a, b]$. Thus the number $M - 1/\mu$, which is smaller than $M$, is the upper bound of $f(x)$ on the interval $[a, b]$. But this contradicts the fact that the number $M$ is the least upper bound of $f(x)$ on the interval $[a, b]$. This contradiction proves that there is a point $x_1 \in [a, b]$ at which $f(x_1) = M$.

We can prove by analogy that the function $f(x)$ attains its greatest lower bound $m$ on $[a, b]$. ■

**Remark 1.** When we have proved that the function $f(x)$ which is continuous on the interval $[a, b]$ attains its least upper bound $M$ and greatest lower bound $m$ on this interval, we can say that the least upper bound is the maximum value and the greatest lower bound is the minimum value of the function $f(x)$ on this interval and formulate Theorem 4.12 as follows: *a function continuous on a closed interval has a maximum and a minimum value on that interval.*

**Remark 2.** The difference of the maximum and minimum values of the continuous function $f(x)$ on the interval $[a, b]$ is the *oscillation* of the continuous function on this interval and is designated as $\omega$: $\omega = M - m$, where

$$M = \max_{[a, b]} f(x), \quad m = \min_{[a, b]} f(x).$$

**5. The concept of the uniform continuity of a function.** Among other properties of a function continuous on an interval we must point out a very important property which is known as uniform continuity. It is widely used in the proof of a number of fundamental theorems.

Assume that $f(x)$ is a function continuous on an interval $X$ and the point $x_0 \in X$. Since the function $f(x)$ is continuous at the point $x_0$, it follows, in accordance with the second definition of continuity, that for any $\varepsilon > 0$ there is $\delta > 0$ such that $|f(x) - f(x_0)| < \varepsilon$ for $|x - x_0| < \delta$. It is clear that $\delta$ depends on $\varepsilon$, but $\delta$ also depends on $x_0$. When $x_0$ varies within the interval being considered (for a constant $\varepsilon$), the number $\delta$ is different for different $x_0$. The steeper the graph of the function $f(x)$ in the neighbourhood of the point $x_0$, the smaller $\delta$ that corresponds to this point (Fig. 59).

Thus, for a given $\varepsilon$, every point $x$ from the interval in question is associated with a certain $\delta > 0$. If there were a finite number of points, then we could choose, from a finite set of the numbers $\delta$, the least positive $\delta$ which would only depend on $\varepsilon$ and would "suit" all $x$. It cannot be done, in general, for an infinite number of points since these points are associated with an infinite number of $\delta$'s among which we can find arbitrarily small ones.

A question arises as to whether there are continuous functions defined on some intervals for which for any $\varepsilon > 0$ we could find $\delta > 0$ which does not depend on $x$, i.e. $\delta$ would be common for all $x$ from

the interval being considered. This leads to the concept of the uniform continuity of a function.

**Definition.** *The function $f(x)$ is uniformly continuous on the interval $X$ is for any $\varepsilon > 0$ there is $\delta > 0$ such that an inequality $|f(x'') -$*

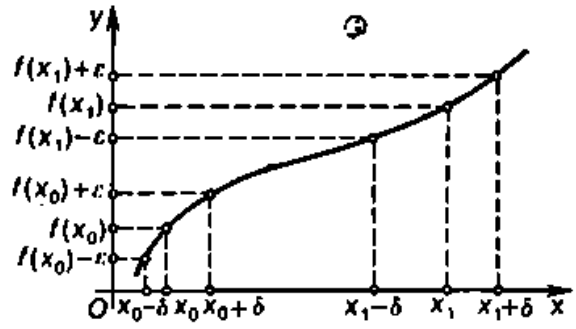



Fig. 59

*$f(x')| < \varepsilon$ holds true for any two points $x'$, $x'' \in X$, which satisfy the inequality $|x'' - x'| < \delta$.*

With the use of logical symbols this definition can be written as

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall x', \quad x'' \in X, \ |x'' - x'| < \delta): \ |f(x'') - f(x')| < \varepsilon.$$

According to the definition, $\delta$ depends only on $\varepsilon$ and is common for all $x'$ and $x''$ from the interval $X$. It is evident from the definition that the uniformly continuous function on $X$ is continuous on this interval.

The following theorem establishes the condition under which a continuous function is also uniformly continuous.

**6. Theorem on the uniform continuity of a function.**

**Theorem 4.13 (Cantor's* theorem).** *If the function $f(x)$ is continuous on the closed interval $[a, b]$, then it is also uniformly continuous on it.*

*Proof.* We shall first prove that if the function $f(x)$ is continuous on $[a, b]$, then for any $\varepsilon > 0$ the closed interval $[a, b]$ can be divided into a finite number of intervals, any two of which either have no points in common or have only one common boundary point and on each of which the inequality $|f(x'') - f(x')| < \varepsilon$ is satisfied for any two points $x'$ and $x''$.

We assume the contrary, i.e. suppose that there is $\varepsilon > 0$ for which such a division of the interval $[a, b]$ is impossible. We bisect $[a, b]$ and choose one of the intervals obtained for which such a division is impossible. We designate it as $[a_1, b_1]$. Then we bisect $[a_1, b_1]$ and choose that of the intervals obtained for which such a division is impossible, and so on. Continuing this process indefinitely, we

* Cantor, Georg Ferdinand Ludwig Philipp (1845-1918), a German mathematician, the founder of the modern set theory.

get a sequence of nested intervals

$$[a, b] \supset [a_1, b_1] \supset [a_2, b_2] \supset \ldots \supset [a_n, b_n] \supset \ldots,$$

which possess the property of being indivisible into a finite number of intervals on each of which the inequality $|f(x'') - f(x')| < \varepsilon$ holds true for any two points $x'$ and $x''$. According to Theorem 2.13 on nested intervals, there is a point $c$ which belongs to all the intervals. Since the function $f(x)$ is continuous at the point $c$, we can find for the number $\varepsilon$ being considered a number $\delta$ such that $|f(x) - f(c)| < \varepsilon/2$ for any $x$ from the $\delta$-neighbourhood of the point $c$. Then for any two points $x'$ and $x''$ from the $\delta$-neighbourhood of the point $c$ the following inequality will be satisfied:

$$|f(x'') - f(x')| = |(f(x'') - f(c)) + (f(c) - f(x'))|$$
$$\leqslant |f(x'') - f(c)| + |f(c) - f(x')| < \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

i.e.

$$|f(x'') - f(x')| < \varepsilon.$$

For a sufficiently large $n$ the $\delta$-neighbourhood of the point $c$ will include the interval $[a_n, b_n]$ and, consequently, the inequality $|f(x'') - f(x')| < \varepsilon$ holds true for any two points $x'$ and $x''$ of this interval, and this contradicts the choice of the sequence of nested intervals.

We shall now prove the theorem. In accordance with what we have just proved, for any $\varepsilon > 0$ there is a division of $[a, b]$ into a finite number of intervals in each of which the difference of any two values of the function $f(x)$ is smaller than $\varepsilon/2$ in absolute value. We denote by $\delta$ the length of the smallest of the intervals resulting from the division and consider any two points $x'$ and $x''$ of the interval $[a, b]$ which are less than $\delta$ distant from each other, i.e. $|x'' - x'| < \delta$. Two cases are possible here: (1) the points $x'$ and $x''$ belong to the same interval resulting from the division; (2) the points $x'$ and $x''$ belong to two neighbouring intervals. In the first case $|f(x'') - f(x')| < \varepsilon/2 < \varepsilon$; in the second case, denoting the boundary point common for the neighbouring intervals by $x_0$, we obtain

$$|f(x'') - f(x')| = |(f(x'') - f(x_0)) + (f(x_0) - f(x'))|$$
$$\leqslant |f(x'') - f(x_0)| + |f(x_0) - f(x')| < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

Thus, for any $\varepsilon > 0$ there is $\delta > 0$ such that an inequality $|f(x'') - f(x')| < \varepsilon$ holds true for any two points $x'$ and $x''$ from the interval $[a, b]$, which satisfy the inequality $|x'' - x'| < \delta$, and this is what we wished to prove. ■

**Corollary.** *Let the function $f(x)$ be continuous on the closed interval $[a, b]$. Then for any $\varepsilon > 0$ there is $\delta > 0$ such that if we arbitrarily divide $[a, b]$ into a finite number of intervals whose lengths are smaller*

*than* $\delta$, *then on each of them the oscillation* $\omega$ *of the function* $f(x)$ *will be smaller than* $\varepsilon$.

*Proof.* Indeed, according to the theorem we have proved, the function $f(x)$ is uniformly continuous on $[a, b]$. Consequently, for any $\varepsilon > 0$ there is $\delta > 0$ such that an inequality $|f(x'') - f(x')| < \varepsilon$ holds true for any points $x'$ and $x''$ from the interval $[a, b]$, which satisfy the inequality $|x'' - x'| < \delta$. We divide arbitrarily the interval $[a, b]$ into a finite number of intervals whose lengths are

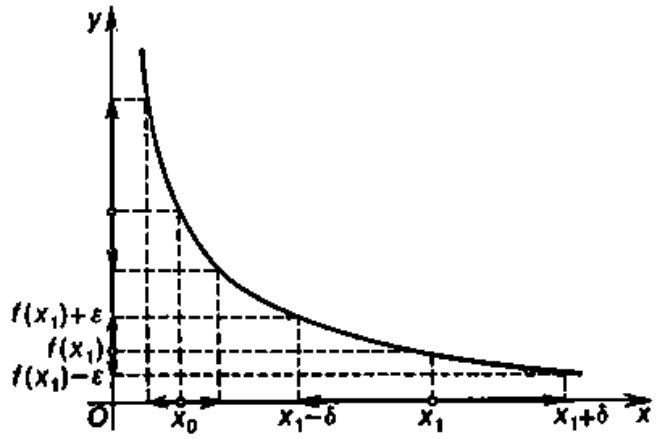



Fig. 60

smaller than the indicated $\delta$. Since the function $f(x)$ is continuous on $[a, b]$, on each of the subintervals we can indicate points $x'$ and $x''$ such that $f(x') = m$ and $f(x'') = M$, where $m$ and $M$ are the greatest lower and the least upper bound of the function $f(x)$ on that subinterval. Since $|x'' - x'| < \delta$, it follows that $|f(x'') - f(x')| < \varepsilon$. But $f(x'') - f(x') = M - m = \omega$ and therefore $\omega < \varepsilon$. ■

**Remark.** The theorem is not true if the closed interval $[a, b]$ is replaced by an open interval or a ray.

**Example.** Let us consider the function $f(x) = 1/x$ on the interval $(0, 1)$. This function is continuous on the interval $(0, 1)$ but is not uniformly continuous on it. This follows from the fact that for any fixed $\varepsilon > 0$, whatever $\delta > 0$ we take, there are always points $x'$ and $x''$, sufficiently close to zero, the distance between which is smaller than $\delta$, and the absolute value of the difference $|f(x'') - f(x')|$ is larger than $\varepsilon$ (Fig. 60).

### 1.4.11. Composite Functions

**Definition.** *If a function* $z = \varphi(x)$ *with a set of values* $Z$ *is defined on an interval* $X$ *and a function* $y = f(z)$ *is defined on a set* $Z$, *then the function* $y = f[\varphi(x)]$ *is a composite function* of* $x$ *and the variable* $z$ *is an intermediate variable of the composite function.*

* Along with the term "composite function" we can use here an equivalent term "superposition of functions".

**Example.** The function $y = \sin x^2$ is a composite function defined throughout the number line since $y = f(z) = \sin z$ and $z = \varphi(x) = x^2$.

**Theorem 4.14.** *Assume that the function $z = \varphi(x)$ is continuous at the point $x_0$ and the function $y = f(z)$ is continuous at the point $z_0 = \varphi(x_0)$. Then the composite function $y = f[\varphi(x)]$ is continuous at the point $x_0$.*

*Proof.* We take from $X$ any sequence of points

$$x_1, x_2, x_3, \ldots, x_n, \ldots,$$

which converges to the point $x_0$. Then, by virtue of the continuity of the function $z = \varphi(x)$ at the point $x_0$, we have $\lim\limits_{n\to\infty} z_n = \lim\limits_{n\to\infty} \varphi(x_n) = \varphi(x_0) = z_0$, i.e. the corresponding sequence of points

$$z_1, z_2, z_3, \ldots, z_n, \ldots$$

converges to the point $z_0$. By virtue of the continuity of the function $f(z)$ at the point $z_0$, we get $\lim\limits_{n\to\infty} f(z_n) = f(z_0)$, i.e. $\lim\limits_{n\to\infty} f[\varphi(x_n)] = f[\varphi(x_0)]$. Consequently, the limit of the function $f[\varphi(x)]$ at the point $x_0$ is equal to its value at this point, and this proves that the composite function $f[\varphi(x)]$ is continuous at the point $x_0$. ■

**Example.** Prove the continuity of the function $y = \sin x^2$ at the point $x = 0$.

*Solution.* The function $z = x^2$ is continuous at the point $x = 0$ and the function $y = \sin z$ is continuous at the point $z = 0$. Therefore, according to the theorem we have proved, the composite function $y = \sin x^2$ is continuous at the point $x = 0$.

### 1.4.12. Inverse Functions

**1. The definition of an inverse function.** We say that the function $f(x)$ does not decrease (does not increase) on the set $X$ if for any $x_1, x_2 \in X$, which satisfy the condition $x_1 < x_2$, the inequality $f(x_1) \leqslant f(x_2)$ ($f(x_1) \geqslant f(x_2)$) holds true.

Nondecreasing and nonincreasing functions taken together are called *monotone functions.*

If for any $x_1, x_2 \in X$, which satisfy the condition $x_1 < x_2$, the inequality $f(x_1) < f(x_2)$ ($f(x_1) > f(x_2)$) holds true, then the function $f(x)$ is *increasing* (*decreasing*) on the set $X$. Increasing and decreasing functions are also called *strictly monotone functions.*

**Examples.**

1. The function $f(x) = \operatorname{sgn} x$ is a nondecreasing function on the entire number line.

2. The function $f(x) = x$ is an increasing function on the entire number line.

We shall introduce now the concept of an inverse function.

**Definition.** *Assume that $X$ and $Y$ are some sets and let a function $f$ be defined, i.e. a set of pairs of numbers $(x, y)$ $(x \in X, y \in Y)$, in which every number $x$ belongs to one and only one pair and every number $y$ belongs to at least one pair. If we make $x$ and $y$ change places in every pair of this set, we get a set of pairs of numbers $(y, x)$ which is known as the inverse of the function $f$.*

We shall designate the inverse function as $x = \varphi(y)$.

Note that the inverse function is, generally speaking, not a function since every number $y$ may belong not to one but to several pairs. Thus, for instance, the inverse $x = y$ of the function $y = x$

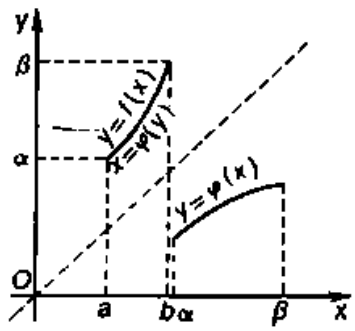



Fig. 61

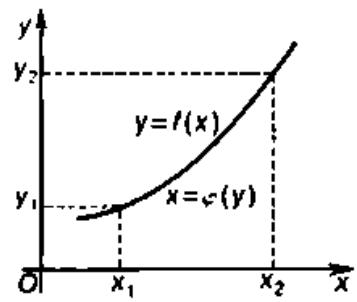



Fig. 62

is single-valued (every number $y$ belongs to one pair), the inverse $x = \pm\sqrt{y}$ of the function $y = x^2$ is double-valued (every number $y$ belongs to two pairs) and the inverse $x = \operatorname{Arcsin} y$ of the function $y = \sin x$ is multivalued (every number $y$ belongs to an infinite number of pairs). From the point of view of geometry, this fact is obvious.

It follows from the definition that if the inverse function is single-valued, i.e. is a function in the ordinary sense, then the range $Y$ of the function $f$ is the domain of definition of the inverse function $\varphi$, and the domain of definition $X$ of the function $f$ is the range of the inverse function $\varphi$. Assume, for instance, that the function $y = f(x)$ is defined on the interval $[a, b]$, the interval $[\alpha, \beta]$ is its range, and every $y \in [\alpha, \beta]$ corresponds to exactly one $x$ from $[a, b]$. Then, according to the definition, a single-valued inverse function $x = \varphi(y)$ is defined on the interval $[\alpha, \beta]$ whose range is the interval $[a, b]$ (Fig. 61).

Thus the function $y = f(x)$ and the inverse function $x = \varphi(y)$ have the same graph. For instance, the function $y = 5x$ and the inverse function $x = 1/5y$ are graphically represented by the same straight line.

If we interchange the $x$-axis and $y$-axis, for which purpose we must rotate the $xy$-plane in space about the bisector of the first quadrant through 180°, then the new position of the graph of the inverse function $x = \varphi(y)$ is the graph of the inverse function $y = \varphi(x)$ (see Fig. 61).

**2. Theorem on the continuity of an inverse function.**

**Theorem 4.15.** *Assume that the function $y = f(x)$ is defined, strictly monotone and continuous on an interval $X$ and let $Y$ be its range. Then, on the set $Y$, the inverse function $x=\varphi(y)$ is single-valued, strictly monotone and continuous.*

*Proof.* Assume, for definiteness, that the function $f(x)$ increases on $X$, i.e. the inequality $y_1 < y_2$ ($y_1 = f(x_1)$, $y_2 = f(x_2)$) holds true for any $x_1, x_2 \in X$, which satisfy the condition $x_1 < x_2$ (Fig. 62).

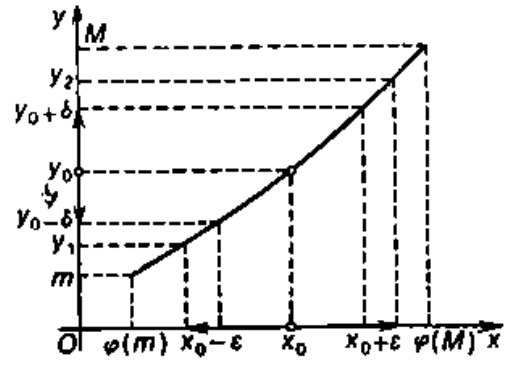



Fig. 63

The inverse function $x=\varphi(y)$ is single-valued since, by virtue of the increase of the function $y = f(x)$ on $X$, the inequality $y_1 = f(x_1) \neq f(x_2) = y_2$ holds true for $x_1 \neq x_2$, and this means that every $y \in Y$ is associated with a single value $x \in X$.

We shall now prove that the inverse function $x = \varphi(y)$ increases on $Y$. Indeed, if $y_1 < y_2$, then $x_1 < x_2$ ($x_1 = \varphi(y_1)$, $x_2 = \varphi(y_2)$) as well since should we have $x_1 \geqslant x_2$, it would follow from the increase of $f(x)$ that $y_1 \geqslant y_2$, and this would contradict the assumption that $y_1 < y_2$. We have thus established that the inverse function $x = \varphi(y)$ is strictly monotone.

We shall finally show that the inverse function $x = \varphi(y)$ is continuous on $Y$. According to the corollary of Theorem 4.10, the set $Y$ is an open interval with endpoints $m$ and $M$, where $m = \inf\limits_X f(x)$ and $M = \sup\limits_X f(x)$. Let $y_0 \in Y$ and $x_0 = \varphi(y_0)$. We shall first consider the case when $m < y_0 < M$ (Fig. 63). In this case the point $x_0$ is evidently an interior point of the interval $X$. We take $\varepsilon > 0$ such that $(x_0 - \varepsilon) \in X$, $(x_0 + \varepsilon) \in X$ and set $y_1 = f(x_0 - \varepsilon)$, $y_2 = f(x_0 + \varepsilon)$. Then, since $f(x)$ is increasing, we get

$$y_1 < y_0 < y_2.$$

We take now $\delta > 0$ such that the inequalities $y_1 \leqslant y_0 - \delta$ and $y_0 + \delta \leqslant y_2$ are satisfied. Then, if $y$ satisfies the inequalities

$$y_0 - \delta < y < y_0 + \delta,$$

we have

$$y_1 < y < y_2,$$

and, consequently, since $\varphi(y)$ is increasing, it follows that

$$\varphi(y_1) < \varphi(y) < \varphi(y_2).$$

Taking into account that $\varphi(y_1) = x_0 - \varepsilon = \varphi(y_0) - \varepsilon$ and $\varphi(y_2) = x_0 + \varepsilon = \varphi(y_0) + \varepsilon$, we get $\varphi(y_0) - \varepsilon < \varphi(y) < \varphi(y_0) + \varepsilon$ under the condition that $y_0 - \delta < y < y_0 + \delta$.

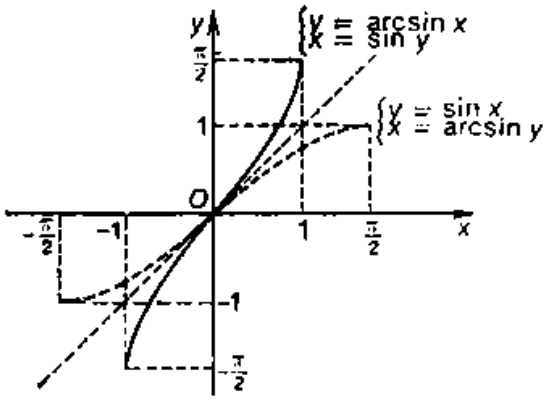


Fig. 64

We have thus proved that for any sufficiently small $\varepsilon > 0$ there is $\delta > 0$ such that the inequality $|\varphi(y) - \varphi(y_0)| < \varepsilon$ holds true for all $y$ which satisfy the inequality $|y - y_0| < \delta$, i.e. the inverse function $\varphi(y)$ is continuous at the point $y_0$. But $y_0$ is an arbitrary point from the interval $(m, M)$. Hence the inverse function $\varphi(y)$ is continuous on $(m, M)$.

If $m \in Y$ or $M \in Y$, then, reasoning by analogy, we can prove that the function $\varphi(y)$ is continuous at the point $m$ from the right and at the point $M$ from the left.

We have thus proved that the inverse function $x = \varphi(y)$ is continuous on $Y$.

In the case when the function $f(x)$ is decreasing the proof of the theorem is similar. ■

**Remark.** If the inverse function $x = \varphi(y)$ is single-valued, then, evidently, the function $y = f(x)$ is an inverse of the function $x = \varphi(y)$. Functions of this kind are also known as *mutually inverse functions*.

**Example.** On the interval $[-\pi/2, \pi/2]$ the function $y = \sin x$ increases, is continuous and its range is the interval $[-1, 1]$. According to Theorem 4.15, there is a continuous increasing inverse function with a set of values $[-\pi/2, \pi/2]$ on the interval $[-1, 1]$. We designate this function as $x = \arcsin y$. Its graph coincides with the graph of the function $y = \sin x$ considered for $-\pi/2 \leqslant x \leqslant \pi/2$ (Fig. 64).

If we now interchange $x$ and $y$, i.e. consider the function $y = \arcsin x$, then we obtain a graph shown in Fig. 64 by the solid line.

Chapter 5

# DIFFERENTIATION

## 1.5.1. The Derivative

**1. Definition of the derivative.** Let the function $y = f(x)$ be defined on a closed interval $X$. We take an arbitrary point $x_0 \in X$ and assign an arbitrary increment $\Delta x$ to the argument $x$ at the point $x_0$, the increment being such that the point $x_0 + \Delta x$ also belongs to $X$. The function gets an increment $\Delta y = f(x_0 + \Delta x) - f(x_0)$.

**Definition.** *The derivative of the function $y = f(x)$ at the point $x_0$ is the limit of the ratio of the increment of the function at this point to the increment of the argument as $\Delta x \to 0$ (provided that this limit exists).*

The symbols $y'(x_0)$ or $f'(x_0)$ are used to denote the derivative of the function $y = f(x)$ at the point $x_0$.

Thus, by definition

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}.$$

If the condition

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = +\infty \quad \left(\text{or } \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = -\infty\right)$$

is fulfilled for a value $x_0$, then we say that at the point $x_0$ the function has an *infinite derivative* of the plus sign (or of the minus sign). As distinct from an infinite derivative, the derivative of a function defined above is sometimes called a *finite derivative*. If the function $f(x)$ has a finite derivative at every point $x \in X$, then we can regard the derivative $f'(x)$ as a function of $x$ which is also defined on $X$.

The definition of a derivative yields a method of its calculation.

**Example.** Find the derivative of the function $f(x) = x^2$ at the point $x = x_0$.

*Solution.* Assigning an increment $\Delta x$ to the argument $x$ at the point $x_0$, we find the corresponding increment of the function:

$$\Delta y = f(x_0 + \Delta x) - f(x_0) = (x_0 + \Delta x)^2 - x_0^2$$
$$= x_0^2 + 2x_0 \Delta x + (\Delta x)^2 - x_0^2 = 2x_0 \Delta x + (\Delta x)^2.$$

We form a ratio $\frac{\Delta y}{\Delta x}$

$$\frac{\Delta y}{\Delta x} = \frac{2x_0 \Delta x + (\Delta x)^2}{\Delta x}$$

and find the limit of this ratio as $\Delta x \to 0$:

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{2x_0 \Delta x + (\Delta x)^2}{\Delta x} = 2x_0.$$

Consequently, the derivative of the function $f(x) = x^2$ at the point $x_0$ is equal to the number $2x_0$. In the symbols we have accepted we can write this as $f'(x_0) = 2x_0$.

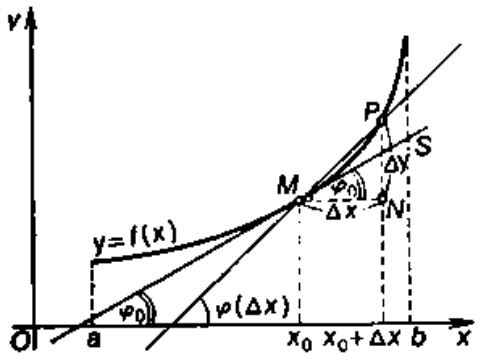



Fig. 65

**2. The geometric meaning of the derivative.** Assume that the function $y = f(x)$ is defined on the open interval $(a, b)$ and the point $M$ on the graph of the function corresponds to the value of the argument $x_0$ and the point $P$ to the value $x_0 + \Delta x$. We draw a straight line through the points $M$ and $P$ and call it a *secant*. Then we denote the angle between the secant and the $x$-axis by $\varphi(\Delta x)$ (Fig. 65). This angle evidently depends on $\Delta x$.

If there is a limit $\lim\limits_{\Delta x \to 0} \varphi(\Delta x) = \varphi_0$, then the straight line with the slope $k = \tan \varphi_0$, which passes through the point $M(x_0, f(x_0))$, is called the *limiting position of the secant MP* as $\Delta x \to 0$ (or as $P \to M$).

**Definition.** *The tangent S to the graph of the function $y = f(x)$ at the point M is the limiting position of the secant MP as $\Delta x \to 0$, or, what is the same, as $P \to M$.*

It follows from the definition that for a tangent to exist it is sufficient that the limit $\lim\limits_{\Delta x \to 0} \varphi(\Delta x) = \varphi_0$ exists, the limit $\varphi_0$ being equal to the angle of inclination of the tangent to the $x$-axis.

We shall prove that if the function $y = f(x)$ has a derivative at the point $x_0$, then there is a tangent to the graph of the function $y = f(x)$ at the point $M(x_0, f(x_0))$, the slope of the tangent (i.e. the tangent of the angle of its inclination to the $x$-axis) being equal to the derivative $f'(x_0)$.

Indeed, from the triangle $MNP$ we get

$$\tan \varphi(\Delta x) = \frac{\Delta y}{\Delta x} = \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}.$$

Hence

$$\varphi(\Delta x) = \arctan \frac{\Delta y}{\Delta x}. \tag{1}$$

We pass to the limit in relation (1) as $\Delta x \to 0$. Since there is a derivative $f'(x_0)$, there is also a limit

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = f'(x_0).$$

From this and from the continuity of the function $\arctan \frac{\Delta y}{\Delta x}$ it follows that there is a limit of the right-hand side of relation (1):

$$\lim_{\Delta x \to 0} \arctan \frac{\Delta y}{\Delta x} = \arctan \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \arctan f'(x_0).$$

Consequently, there is also a limit of the left-hand side of relation (1). Thus we have

$$\lim_{\Delta x \to 0} \varphi(\Delta x) = \arctan f'(x_0).$$

And this precisely means that there is a limiting position of the secant $MP$, i.e. there is a tangent to the graph of the function $y = f(x)$ at the point $M(x_0, f(x_0))$, the angle of inclination $\varphi_0$ of that tangent to the $x$-axis being equal to $\arctan f'(x_0)$, and this means that the slope of the tangent $\tan \varphi_0 = f'(x_0)$, and this is what we wished to prove.

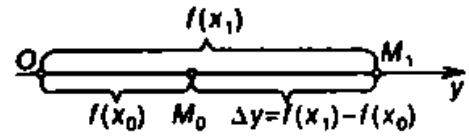


Fig. 66

Thus the derivative of the function $y = f(x)$ at the point $x_0$ is equal to the slope of the tangent to the graph of the function $y = f(x)$ at the point $M(x_0, f(x_0))$.

**3. The physical meaning of the derivative.** We assume that the function $y = f(x)$ describes the law of motion of a particle $M$ along a straight line, i.e. $y = f(x)$ is the path traversed by the particle $M$ from the reference point in time $x$.

Then, in time $x_0$ the particle covered the distance $y = f(x_0)$ and in time $x_1$ the distance $y = f(x_1)$. In time $\Delta x = x_1 - x_0$ the particle $M$ will cover the distance $\Delta y = f(x_1) - f(x_0) = f(x_0 + \Delta x) - f(x_0)$ (Fig. 66). The ratio $\frac{\Delta y}{\Delta x}$ is known as the *average velocity* of motion ($v_{av}$) in the time $\Delta x$, and the limit of the ratio $\frac{\Delta y}{\Delta x}$, as $\Delta x \to 0$, defines the *instantaneous velocity* of the particle at the moment $x_0$ ($v_{ins}$).

It is convenient to use the concept of the velocity, taken from physics, for investigating the behaviour of an arbitrary function. Whatever the dependence that the function $y = f(x)$ shows, the ratio $\frac{\Delta y}{\Delta x}$ is the average velocity of variation of $y$ relative to the variation of $x$, and $y'(x_0)$ is the instantaneous velocity of variation of $y$ for $x = x_0$.

The significance of the derivative is that when we study some processes and phenomena of nature, we can use it to estimate the rate of variation of some related quantities.

**4. The right-hand and left-hand derivatives.** Using the notions of the right-hand and left-hand limits of a function, we shall introduce the concepts of the right-hand and left-hand derivatives of the function $y = f(x)$ at the point $x_0$.

**Definition.** *The right-hand (left-hand) derivative of the function $y = f(x)$ at the point $x_0$ is the right-hand (left-hand) limit of the ratio $\frac{\Delta y}{\Delta x}$ as $\Delta x \to 0$ (provided that this limit exists).*

The notation is

$$f'_+(x_0) = \lim_{\Delta x \to 0+} \frac{\Delta y}{\Delta x} \quad \left(f'_-(x_0) = \lim_{\Delta x \to 0-} \frac{\Delta y}{\Delta x}\right).$$

If the function $f(x)$ has a derivative at the point $x_0$, then it has a right-hand and a left-hand derivative at this point, which coincide. At the same time, there are functions which possess a right-hand and a left-hand derivative at the point $x_0$ but do not have a derivative at this point. Thus, for instance, the function $f(x) = |x|$ has a right-hand derivative equal to $f'_+(0) = \lim\limits_{\Delta x \to 0+} \frac{\Delta y}{\Delta x} = 1$ ($\Delta y = \Delta x$ for $x \geqslant 0$) and a left-hand derivative equal to $f'_-(0) = \lim\limits_{\Delta x \to 0-} \frac{\Delta y}{\Delta x} = -1$ ($\Delta y = -\Delta x$ for $x < 0$) at the point $x = 0$ but does not have a derivative at this point since $f'_+(0) \neq f'_-(0)$.

### 1.5.2. The Differentiability of a Function

**1. The concept of the differentiability of a function at a given point.**

**Definition.** *The function $y = f(x)$ is differentiable at the point $x_0$ if its increment $\Delta y$ at this point can be represented as*

$$\Delta y = A\,\Delta x + \alpha(\Delta x)\,\Delta x, \tag{1}$$

*where $A$ is a number independent of $\Delta x$ and $\alpha(\Delta x)$ is a function of the argument $\Delta x$ which is infinitesimal as $\Delta x \to 0$, i.e. $\lim\limits_{\Delta x \to 0} \alpha(\Delta x) = 0$.*

Let us establish the relationship between the differentiability of a function at a point and the existence of a derivative at that point.

**Theorem 5.1.** *For the function $y = f(x)$ to be differentiable at the point $x_0$, it is necessary and sufficient that it should possess a finite derivative at this point.*

*Proof.* **Necessity.** Assume that the function $y = f(x)$ is differentiable at the point $x_0$, i.e. its increment at this point can be represented in form (1): $\Delta y = A\,\Delta x + \alpha(\Delta x)\,\Delta x$. Dividing this equality

by $\Delta x$ (for $\Delta x \neq 0$), we obtain

$$\frac{\Delta y}{\Delta x} = A + \alpha(\Delta x).$$

Passing to the limit as $\Delta x \to 0$, we have

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} (A + \alpha(\Delta x)) = A.$$

From this is follows that at the point $x_0$ the derivative exists and is equal to $A$: $f'(x_0) = A$.

**Sufficiency.** Assume that there is a finite derivative $f'(x_0)$, i.e. $\lim\limits_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = f'(x_0)$. Let $f'(x_0) = A$. Then the function $\alpha(\Delta x) = \frac{\Delta y}{\Delta x} - A$ is an infinitesimal as $\Delta x \to 0$ (see Theorem 4.5) From the last relation we have

$$\Delta y = A\,\Delta x + \alpha(\Delta x),$$

where $\lim\limits_{\Delta x \to 0} \alpha(\Delta x) = 0$. We have got representation (1) and thus proved that the function $y = f(x)$ is differentiable at the point $x_0$. ■

Thus, for a function of one variable, the concepts of the differentiability and the existence of a derivative are equivalent, and therefore the operation of finding a derivative is often called differentiation.

**Remark.** The function $\alpha(\Delta x) = \frac{\Delta y}{\Delta x} - A$ we have introduced when we proved the sufficiency is not defined for $\Delta x = 0$. Consequently, expression (1) obtained for $\Delta y$ is not defined for $\Delta x = 0$ either. If we arbitrarily define $\alpha(0)$, then relation (1) will be valid for $\Delta x = 0$ as well. For what follows, it is advisable to agree that in expression (1) the function $\alpha(\Delta x)$ is defined by continuity for $\Delta x = 0$, i.e. $\alpha(0) = \lim\limits_{\Delta x \to 0} \alpha(\Delta x) = 0$.

**2. The relationship between the concepts of differentiability and continuity.**

**Theorem 5.2.** *If the function $y = f(x)$ is differentiable at a given point $x_0$, then it is continuous at this point.*

*Proof.* Since the function $y = f(x)$ is differentiable at the point $x_0$, its increment at this point can be represented by relation (1). Then, passing to the limit as $\Delta x \to 0$, we obtain

$$\lim_{\Delta x \to 0} \Delta y = A \lim_{\Delta x \to 0} \Delta x + \lim_{\Delta x \to 0} \alpha(\Delta x) \lim_{\Delta x \to 0} \Delta x = 0,$$

and this signifies the continuity of the function $y = f(x)$ at the point $x_0$ according to the third definition of the continuity of a function at a point. ■

**Remark.** The converse statement is not true. A function may be continuous at a point but not differentiable, i.e. may have no derivative at that point.

The function $f(x) = |x|$ may serve as an example of such a function. It is continuous at a point $x = 0$ but, as was shown in Item 4 of 1.5.1, has no derivative at this point, i.e. is not differentiable.

If the function $f(x)$ has a derivative at every point of a certain closed interval (is differentiable at every point of the interval), then we say that the function $f(x)$ is differentiable on that interval.

### 1.5.3. The Differential

**1. The definition and the geometric meaning of the differential.** Assume that the function $y = f(x)$ is differentiable at the point $x_0$, i.e. its increment $\Delta y$ at this point can be written as the sum of two terms:

$$\Delta y = A\,\Delta x + \alpha(\Delta x)\,\Delta x,$$

where $\lim\limits_{\Delta x \to 0} \alpha(\Delta x) = 0$. As $\Delta x \to 0$, the term $A\,\Delta x$ is an infinitesimal of the same order as $\Delta x$ (for $A \neq 0$). It is linear relative to $\Delta x$. As $\Delta x \to 0$, the term $\alpha(\Delta x)\,\Delta x$ is an infinitesimal of a higher order than $\Delta x$ $\left(\lim\limits_{\Delta x \to 0} \frac{\alpha(\Delta x)\,\Delta x}{\Delta x} = 0\right)$.

Thus, for $A \neq 0$, the first term is the principal part of the increment of the function $y = f(x)$.

**Definition.** *The differential of the function $y = f(x)$ at the point $x_0$ is the principal, linear relative to $\Delta x$, part of the increment of the function at this point:*

$$dy = A\,\Delta x. \tag{1}$$

If $A = 0$, then $A\,\Delta x = 0$, and therefore the term $A\,\Delta x$ is no longer the principal part of the increment $\Delta y$ since the term $\alpha(\Delta x)\,\Delta x$ is, in general, different from zero. However, by definition, in this case as well we set the differential of the function at the point $x_0$ equal to $A\,\Delta x$, i.e. here $dy = 0$.

Bearing Theorem 5.1 in mind, i.e. taking into account that $A = f'(x_0)$, we can write formula (1) in the form

$$dy = f'(x_0)\,\Delta x. \tag{2}$$

Let $f(x) = x$. Then, by formula (2),

$$dy = dx = (x_0)'\,\Delta x = \left(\lim_{\Delta x \to 0} \frac{(x_0 + \Delta x) - x_0}{\Delta x}\right)\Delta x$$

$$= \left(\lim_{\Delta x \to 0} \frac{\Delta x}{\Delta x}\right)\Delta x = 1 \cdot \Delta x = \Delta x.$$

The differential of the independent variable $x$ is the increment $dx = \Delta x$ of this variable. Now relation (2) assumes the form

$$dy = f'(x_0)\,dx. \tag{3}$$

Note that using relation (3), we can calculate the derivative $f'(x_0)$ as the ratio of the differential $dy$ of the function to the differential $dx$ of the independent variable, i.e.

$$f'(x_0) = \frac{dy}{dx}.$$

The differential of a function has a geometric meaning. Assume that a point $M$ on the graph of the function $y = f(x)$ corresponds to the value of the argument $x_0$, a point $P$, to the value of the argument $x_0 + \Delta x$, a straight line $MS$ is a tangent of the graph of $y = f(x)$ at the point $M$, and $\alpha$ is the angle between the tangent and the $x$-axis. Assume, furthermore, that $MN \parallel Ox$, $PN \parallel Oy$ and $Q$ is the point of intersection of the tangent $MS$ and the straight line $PN$ (Fig. 67). Then the increment $\Delta y$ of the function is equal to the value of the line segment $NP$. At the same time, from the right triangle $MNQ$ we get $NQ = \tan\alpha\,\Delta x = f'(x_0)\,\Delta x = dy$, i.e. the differential $dy$ of the function is equal to the value of the line segment $NQ$. In terms of geometry, the values of the segments $NP$ and $NQ$ are different.

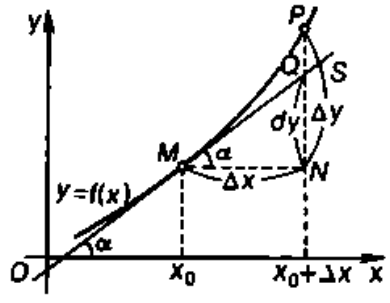


Fig. 67

Thus the differential $dy$ of the function $y = f(x)$ at the point $x_0$ is equal to the increment of the "ordinate of the tangent" $MS$ to the graph of this function at the point $M(x_0, f(x_0))$, and the increment $\Delta y$ of the function is the increment of the "ordinate of the function" $y = f(x)$ itself at the point $x_0$ corresponding to the increment $\Delta x$ of the argument.

**2. Approximation by means of a differential.** It follows from the definition of the differential that it is in a linear relationship with $\Delta x$ and is the principal part of the increment $\Delta y$ of the function, but $\Delta y$ itself is in a more complicated relationship with $\Delta x$. For instance, if $f(x) = x^3$, then $\Delta y = (x_0 + \Delta x)^3 - x_0^3 = 3x_0^2\,\Delta x + 3x_0(\Delta x)^2 + (\Delta x)^3$, whereas

$$dy = f'(x_0)\,\Delta x = \left(\lim_{\Delta x \to 0} \frac{(x_0 + \Delta x)^3 - x_0^3}{2}\right)\Delta x = 3x_0^2\,\Delta x.$$

In many problems the increment of a function at a given point can be approximated by the differential of the function at that point,

i.e.

$$\Delta y \cong dy.$$

The absolute error resulting from such an approximation is equal to $|\Delta y - dy|$ and, as $\Delta x \to 0$, is an infinitesimal of a higher order than $\Delta x$.

**Example.** We shall show that if $\alpha$ is small, then we can use an approximate formula

$$\sqrt{1+\alpha} \cong 1+\frac{\alpha}{2}.$$

*Solution.* Let us consider a function $f(x) = \sqrt{x}$. For small $\Delta x$ we have

$$\Delta y = \sqrt{x_0+\Delta x} - \sqrt{x_0} \cong dy,$$

or

$$\sqrt{x_0+\Delta x} - \sqrt{x_0}$$

$$\cong (\sqrt{x})'|_{x=x_0}\,\Delta x = \left(\lim_{\Delta x\to 0}\frac{\sqrt{x_0+\Delta x}-\sqrt{x_0}}{\Delta x}\right)\Delta x$$

$$= \left(\lim_{\Delta x\to 0}\frac{(x_0+\Delta x)-x_0}{\Delta x(\sqrt{x_0+\Delta x}+\sqrt{x_0})}\right)\Delta x = \frac{1}{2\sqrt{x_0}}\,\Delta x,$$

whence, setting $x_0 = 1$ and $\Delta x = \alpha$, we get

$$\sqrt{1+\alpha} \cong 1+\frac{\alpha}{2}.$$

In particular, $\sqrt{1.0003} \cong 1.00015$ for $\alpha = 0.0003$.

We shall establish now the rules of differentiation and calculation of the derivatives of the simplest elementary functions. It should only be noted that when deriving formulas and calculating derivatives in practical applications, we usually write $x$ rather than $x_0$ but consider $x$ to be fixed.

### 1.5.4. The Rules of Differentiating the Sum, the Difference, the Product and the Quotient

**Theorem 5.3.** *If the functions $u = u(x)$ and $v = v(x)$ are differentiable at the point $x$, then the sum, the difference, the product and the quotient of these functions (the quotient under the condition that $v(x) \neq 0$) are also differentiable at this point, and the following formulas hold true:*

$$(u \pm v)' = u' \pm v', \quad (u\cdot v)' = u'v + uv', \quad \left(\frac{u}{v}\right)' = \frac{u'v-uv'}{v^2}. \qquad (1)$$

*Proof.* To derive formulas (1), we shall use the definition of the derivative, the equation $f(x+\Delta x) = f(x) + \Delta y$ and Theorem 4.3

Then we obtain

$$(u \pm v)' = \lim_{\Delta x \to 0} \frac{[u(x+\Delta x) \pm v(x+\Delta x)] - [u(x) \pm v(x)]}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \left[ \frac{u(x+\Delta x) - u(x)}{\Delta x} \pm \frac{v(x+\Delta x) - v(x)}{\Delta x} \right]$$

$$= \lim_{\Delta x \to 0} \frac{u(x+\Delta x) - u(x)}{\Delta x} \pm \lim_{\Delta x \to 0} \frac{v(x+\Delta x) - v(x)}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} \pm \lim_{\Delta x \to 0} \frac{\Delta v}{\Delta x} = u' \pm v',$$

$$(uv)' = \lim_{\Delta x \to 0} \frac{u(x+\Delta x)\, v(x+\Delta x) - u(x)\, v(x)}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{[u(x) + \Delta u]\, [v(x) + \Delta v] - u(x)\, v(x)}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{u(x)\, v(x) + v(x)\, \Delta u + u(x)\, \Delta v + \Delta u\, \Delta v - u(x)\, v(x)}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \left[ v(x) \frac{\Delta u}{\Delta x} + u(x) \frac{\Delta v}{\Delta x} + \Delta v \frac{\Delta u}{\Delta x} \right]$$

$$= v \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} + u \lim_{\Delta x \to 0} \frac{\Delta v}{\Delta x} + \lim_{\Delta x \to 0} \Delta v \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x}$$

$$= vu' + uv' + 0 \cdot u' = u'v + uv',$$

since $\lim\limits_{\Delta x \to 0} \Delta v = 0$ and the factors $u$ and $v$ are independent of $\Delta x$,

$$\left(\frac{u}{v}\right)' = \lim_{\Delta x \to 0} \frac{\dfrac{u(x+\Delta x)}{v(x+\Delta x)} - \dfrac{u(x)}{v(x)}}{\Delta x} = \lim_{\Delta x \to 0} \frac{u(x+\Delta x)\, v(x) - u(x)\, v(x+\Delta x)}{v(x+\Delta x)\, \Delta x\, v(x)}$$

$$= \lim_{\Delta x \to 0} \frac{[u(x) + \Delta u]\, v(x) - u(x)\, [v(x) + \Delta v]}{v(x)\, \Delta x\, [v(x) + \Delta v]} = \lim_{\Delta x \to 0} \frac{uv + v\, \Delta u - uv - u\, \Delta v}{v\, \Delta x\, (v + \Delta v)}$$

$$= \lim_{\Delta x \to 0} \frac{v \dfrac{\Delta u}{\Delta x} - u \dfrac{\Delta v}{\Delta x}}{v^2 + v\, \Delta v} = \frac{v \lim\limits_{\Delta x \to 0} \dfrac{\Delta u}{\Delta x} - u \lim\limits_{\Delta x \to 0} \dfrac{\Delta v}{\Delta x}}{v^2 + v \lim\limits_{\Delta x \to 0} \Delta v} = \frac{u'v - uv'}{v^2}. \quad \blacksquare$$

### 1.5.5. Calculating the Derivatives of a Constant Function and a Power Function, of Trigonometric Functions and Logarithmic Functions

**1. The derivative of a constant function.** The derivative of the function $y = f(x) = C$, where $C$ is a constant number, is expressed by the formula

$$y' = 0.$$

*Proof*. For any $x$ and $\Delta x$ we have $f(x+\Delta x)=C$ and $\Delta y = f(x+\Delta x)-f(x)=0$. Hence $\frac{\Delta y}{\Delta x}=0$ for any $\Delta x \neq 0$ and, consequently,

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = 0. \blacksquare$$

**2. The derivative of a power function.** The derivative of the function $y = x^n$ whose exponent $n$ is a positive integer is expressed by the formula

$$y' = n \cdot x^{n-1}.$$

*Proof*. Using the binomial formula, we can write

$$\Delta y = (x+\Delta x)^n - x^n$$
$$= \left[x^n + nx^{n-1}\Delta x + \frac{n(n-1)}{2!}x^{n-2}(\Delta x)^2 + \ldots + (\Delta x)^n\right] - x^n$$
$$= nx^{n-1}\Delta x + \frac{n(n-1)}{2!}x^{n-2}(\Delta x)^2 + \ldots + (\Delta x)^n.$$

Thus, for $\Delta x \neq 0$, we have

$$\frac{\Delta y}{\Delta x} = nx^{n-1} + \frac{n(n+1)}{2!}x^{n-2}\Delta x + \ldots + (\Delta x)^{n-1}.$$

Since $\lim\limits_{\Delta x \to 0} \Delta x = 0$, $\lim\limits_{\Delta x \to 0} (\Delta x)^2 = 0$, $\ldots$, $\lim\limits_{\Delta x \to 0} (\Delta x)^{n-1} = 0$, it follows that

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = nx^{n-1}. \blacksquare$$

**Remark.** We shall consider the case of a power function whose exponent is a real number in Item 2 of 1.5.9.

**3. The derivatives of trigonometric functions.**

(1) The derivative of the function $y = \sin x$ is expressed by the formula

$$y' = \cos x.$$

*Proof*. We have

$$\Delta y = \sin(x+\Delta x) - \sin x = 2\sin(\Delta x/2)\cos(x+\Delta x/2).$$

Thus, for $\Delta x \neq 0$, we have

$$\frac{\Delta y}{\Delta x} = \frac{2\sin(\Delta x/2)\cos(x+\Delta x/2)}{\Delta x} = \frac{\sin(\Delta x/2)}{\Delta x/2}\cos(x+\Delta x/2).$$

Since $\lim\limits_{\Delta x \to 0} \frac{\sin(\Delta x/2)}{\Delta x/2} = 1$ (the first remarkable limit) and $\lim\limits_{\Delta x \to 0} \cos\left(x + \frac{\Delta x}{2}\right) = \cos x$ by virtue of the continuity of the

function $\cos x$, we have

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \cos x. \blacksquare$$

(2) The derivative of the function $y = \cos x$ is expressed by the formula

$$y' = -\sin x.$$

*Proof.* We have

$$\Delta y = \cos(x + \Delta x) - \cos x = -2 \sin(\Delta x/2) \sin(x + \Delta x/2).$$

Thus, for $\Delta x \neq 0$, we have

$$\frac{\Delta y}{\Delta x} = -\frac{2 \sin(\Delta x/2) \sin(x + \Delta x/2)}{\Delta x} = -\frac{\sin(\Delta x/2)}{\Delta x/2} \sin(x + \Delta x/2).$$

Since $\lim\limits_{\Delta x \to 0} \sin\left(x + \frac{\Delta x}{2}\right) = \sin x$ by virtue of the continuity of the function $\sin x$, we have

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = -\sin x. \blacksquare$$

(3) The derivative of the function $y = \tan x$ is expressed by the formula

$$y' = \frac{1}{\cos^2 x} \quad \left(x \neq \frac{\pi}{2} + n\pi\right).$$

*Proof.* Since $\tan x = \sin x/\cos x$, we find, in accordance with Theorem 5.3, that

$$y' = \frac{(\sin x)' \cos x - \sin x (\cos x)'}{\cos^2 x}$$

$$= \frac{\cos x \cos x - \sin x (-\sin x)}{\cos^2 x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x}$$

and, consequently,

$$y' = \frac{1}{\cos^2 x}. \blacksquare$$

(4) The derivative of the function $y = \cot x$ is expressed by the formula

$$y' = -\frac{1}{\sin^2 x} \quad (x \neq n\pi).$$

*Proof.* Since $\cot x = \cos x/\sin x$, we have, by analogy with the preceding case,

$$y' = \frac{(\cos x)' \sin x - \cos x (\sin x)'}{\sin^2 x}$$

$$= \frac{(-\sin x) \sin x - \cos x \cos x}{\sin^2 x} = -\frac{\sin^2 x + \cos^2 x}{\sin^2 x}$$

and, consequently,

$$y' = -\frac{1}{\sin^2 x}. \blacksquare$$

**4. The derivative of a logarithmic function.** The derivative of the function $y = \log_a x$ $(0 < a \neq 1)$ is expressed by the formula

$$y' = \frac{1}{x} \log_a e = \frac{1}{x \ln a}.$$

*Proof.* We have

$$\Delta y = \log_a (x + \Delta x) - \log_a x = \log_a \frac{x + \Delta x}{x} = \log_a \left(1 + \frac{\Delta x}{x}\right).$$

Thus, for $\Delta x \neq 0$, we have

$$\frac{\Delta y}{\Delta x} = \frac{1}{\Delta x} \log_a \left(1 + \frac{\Delta x}{x}\right) = \frac{1}{x} \frac{x}{\Delta x} \log_a \left(1 + \frac{\Delta x}{x}\right),$$

or

$$\frac{\Delta y}{\Delta x} = \frac{1}{x} \log_a \left[\left(1 + \frac{\Delta x}{x}\right)^{x/\Delta x}\right].$$

Setting $x/\Delta x = h$, we have $\lim\limits_{\Delta x \to 0} \left(1 + \frac{\Delta x}{x}\right)^{x/\Delta x} = \lim\limits_{h \to \infty} \left(1 + \frac{1}{h}\right)^h = e$ (the second remarkable limit), and, since the logarithmic function is continuous, it follows that

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \frac{1}{x} \log_a \left[\lim_{\Delta x \to 0} \left(1 + \frac{\Delta x}{x}\right)^{x/\Delta x}\right]$$
$$= \frac{1}{x} \log_a e = \frac{1}{x \ln a}. \blacksquare$$

**Corollary.** *If* $y = \log_e x = \ln x$, *it follows that* $y' = (\ln x)' = 1/x$.

### 1.5.6. Theorem on the Derivative of an Inverse Function

Assume that the function $y = f(x)$ satisfies the hypothesis of Theorem 4.15 on an inverse function and the function $x = \varphi(y)$ is its inverse. Then there holds the following theorem.

**Theorem 5.4.** *If the function* $y = f(x)$ *has a derivative* $f'(x_0) \neq 0$ *at the point* $x_0$, *then its inverse* $x = \varphi(y)$ *also has a derivative at the corresponding point* $y_0 = f(x_0)$, *and*

$$\varphi'(y_0) = \frac{1}{f'(x_0)}.$$

*Proof.* We assign an increment $\Delta y \neq 0$ to the argument $y$ of the inverse function $x = \varphi(y)$ at the point $y_0$. The function $x = \varphi(y)$

gets an increment $\Delta x$, and $\Delta x \neq 0$ by virtue of the increase (or decrease) of the inverse function. Consequently, we can write

$$\frac{\Delta x}{\Delta y} = \frac{1}{\Delta y/\Delta x}.$$

In this relation we pass to the limit as $\Delta y \to 0$. Since the inverse function $x = \varphi(y)$ is continuous at the point $y_0$ (see Theorem 4.15), it follows that $\Delta x \to 0$ as $\Delta y \to 0$. But as $\Delta x \to 0$, the limit of the right-hand side of the relation exists and is equal to $1/f'(x_0)$. Consequently, there is also a limit of the left-hand side of the relation which, by definition, is equal to $\varphi'(y_0)$. Thus we obtain

$$\varphi'(y_0) = \frac{1}{f'(x_0)}. \blacksquare \qquad (1)$$

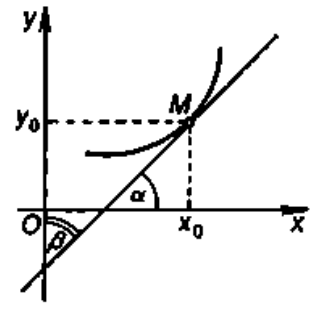


Fig. 68

The theorem we have proved has a simple geometric meaning. Let us consider the graph of the function $y = f(x)$ (or that of the inverse function $x = \varphi(y)$) in some neighbourhood of the point $x_0$. Let the point $x_0$ on this graph be associated with a point $M$ (Fig. 68). As is known, the derivative $f'(x_0)$ is equal to the tangent of the angle of inclination $\alpha$ of the tangent line, which passes through the point $M$, to the $x$-axis. The derivative of the inverse function $\varphi'(y_0)$ is equal to the tangent of the angle of inclination $\beta$ of the same tangent line to the $y$-axis. Since the sum of the angles $\alpha$ and $\beta$ is $\pi/2$, formula (1) expresses the obvious fact:

$$\varphi'(y_0) = \tan\beta = \frac{1}{\cot\beta} = \frac{1}{\cot(\pi/2-\alpha)} = \frac{1}{\tan\alpha} = \frac{1}{f'(x_0)}.$$

### 1.5.7. Calculating the Derivatives of an Exponential Function and of Inverse Trigonometric Functions

Using Theorem 5.4 proved above, we shall continue with the calculations of the simplest elementary functions.

**1. The derivative of an exponential function.** The derivative of the function $y = a^x$ $(0 < a \neq 1)$ is expressed by the formula

$$y' = a^x \ln a.$$

*Proof.* The exponential function $y = a^x$ is the inverse of the logarithmic function $x = \log_a y$. Since

$$x'(y) = \frac{1}{y}\log_a e,$$

we get, by virtue of Theorem 5.4 on the derivative of an inverse function and the familiar relation $\log_a b = 1/\log_b a$, a relation

$$y'(x) = \frac{1}{x'(y)} = \frac{y}{\log_a e} = a^x \ln a. \blacksquare$$

**Corollary.** *If* $y = e^x$, *then* $y' = (e^x)' = e^x$.

**2. The derivatives of inverse trigonometric functions.**

(1) The derivative of the function $y = \arcsin x$ is expressed by the formula

$$y' = \frac{1}{\sqrt{1-x^2}} \quad (|x| < 1).$$

*Proof.* The function $y = \arcsin x$ is the inverse of the function $x = \sin y$. Since $x'(y) = \cos y$, it follows, according to Theorem 5.4 on the derivative of an inverse function, that

$$y'(x) = \frac{1}{x'(y)} = \frac{1}{\cos y} = \frac{1}{\sqrt{1-\sin^2 y}}.$$

The root is taken with the plus sign since $\cos y$ is positive on the interval $-\pi/2 < y < \pi/2$. Taking into account that $\sin y = x$, we finally have

$$y'(x) = \frac{1}{\sqrt{1-x^2}}. \blacksquare$$

(2) The derivative of the function $y = \arccos x$ is expressed by the formula

$$y'(x) = -\frac{1}{\sqrt{1-x^2}}.$$

The proof is similar to the preceding one.

(3) The derivative of the function $y = \arctan x$ is expressed by the formula

$$y' = \frac{1}{1+x^2}.$$

*Proof.* The function $y = \arctan x$ is the inverse of the function $x = \tan y$.

Since $x'(y) = 1/\cos^2 y$, it follows that $y'(x) = 1/x'(y) = \cos^2 y$. But $1/\cos^2 y = 1 + \tan^2 y + 1 + x^2$ and, consequently,

$$y'(x) = \frac{1}{1+x^2}. \blacksquare$$

(4) The derivative of the function $y = \operatorname{arccot} x$ is expressed by the formula

$$y' = -\frac{1}{1+x^2}.$$

The proof is similar to the preceding one.

### 1.5.8. The Rule of Differentiating a Composite Function

**Theorem 5.5.** *If the function $x = \varphi(t)$ has a derivative at the point $t_0$ and the function $y = f(x)$ has a derivative at the corresponding point $x_0 = \varphi(t_0)$, then the composite function $f[\varphi(t)]$ has a derivative at the point $t_0$ and the following formula holds true:*

$$y'(t_0) = f'(x_0)\,\varphi'(t_0). \tag{1}$$

*Proof.* Since the function $y = f(x)$ is differentiable at the point $x_0$, the increment of this function at the point $x_0$ can be written as

$$\Delta y = f'(x_0)\,\Delta x + \alpha(\Delta x)\,\Delta x, \tag{2}$$

where $\lim\limits_{\Delta x \to 0} \alpha(\Delta x) = 0$. We divide relation (2) by $\Delta t$ and get

$$\frac{\Delta y}{\Delta t} = f'(x_0)\frac{\Delta x}{\Delta t} + \alpha(\Delta x)\frac{\Delta x}{\Delta t}. \tag{3}$$

Relation (3) is valid for any sufficiently small $\Delta x$. We take $\Delta x$ equal to the increment of the function $x = \varphi(t)$, which corresponds to the increment $\Delta t$ of the argument $t$ at the point $t_0$, and make $\Delta t$ tend to zero in this relation. Since, by the hypothesis, the function $x = \varphi(t)$ has a derivative at the point $t_0$, it is continuous at this point. Consequently, according to the third definition of the continuity of a function at a point, $\Delta x \to 0$ as $\Delta t \to 0$. But then $\alpha(\Delta x)$ also tends to zero, i.e. we have

$$\lim_{\Delta t \to 0}\left(\alpha(\Delta x)\frac{\Delta x}{\Delta t}\right) = \lim_{\Delta t \to 0}\alpha(\Delta x)\lim_{\Delta t \to 0}\frac{\Delta x}{\Delta t} = 0\cdot\varphi'(t_0) = 0. \tag{4}$$

By virtue of relation (4), there is a limit of the right-hand side of relation (3) as $\Delta t \to 0$, and this limit is equal to $f'(x_0)\,\varphi'(t_0)$. Hence there is also a limit of the left-hand side of relation (3) as $\Delta t \to 0$, which, by the definition of the derivative, is equal to the derivative of the composite function $y = f[\varphi(t)]$ at the point $t_0$. We have thus proved the differentiability of the composite function and established formula (1). ■

**Remark 1.** In this theorem we considered a composite function where $y$ depends on $t$ involving an intermediate variable $x$. A more complicated relationship is also possible, involving two, three or more intermediate variables, but the rule of differentiation remains the same.

Thus, for instance, if $y = f(x)$, where $x = \varphi(u)$, and $u = \psi(v)$ and $v = \chi(t)$, then the derivative $y'(t)$ must be calculated by the formula

$$y'(t) = f'(x)\,\varphi'(u)\,\psi'(v)\,\chi'(t). \tag{5}$$

**Example 1.** Calculate the derivative of the function $y = e^{\arctan x}$.

*Solution.* This function can be represented as $y = e^u$, where $u = \arctan x$. Then, from formula (1), we get

$$y'(x) = y'(u)\,u'(x) = e^u \frac{1}{1+x^2}.$$

Replacing $u$ by $\arctan x$, we finally obtain

$$y' = e^{\arctan x} \frac{1}{1+x^2}.$$

**Example 2.** Calculate the derivative of the function $y = \tan^2\sqrt{x^2+1}$.

*Solution.* This function can be represented as $y = u^2$, where $u = \tan v$, and $v = \sqrt{w}$ and $w = x^2 + 1$. Using formula (5), we obtain

$$y'(x) = y'(u)\,u'(v)\,v'(w)\,w'(x) = (u^2)'\,(\tan v)'\left(\sqrt{w}\right)'(x^2+1)'$$
$$= 2u\sec^2 v \frac{1}{2\sqrt{w}}\,2x = \frac{2x\tan\sqrt{x^2+1}\,\sec^2\sqrt{x^2+1}}{\sqrt{x^2+1}}.$$

**Remark 2.** Sometimes we have to calculate a derivative proceeding directly from its definition. Let us find, for example, the derivative of the function

$$f(x) = \begin{cases} x^2\sin(1/x) & \text{for } x \neq 0, \\ 0 & \text{for } x = 0. \end{cases}$$

For $x \neq 0$ the derivative must be calculated by the formulas and rules of differentiation:

$$f'(x) = \left(x^2\sin\frac{1}{x}\right)' = (x^2)'\sin\frac{1}{x} + x^2\left(\sin\frac{1}{x}\right)'\left(\frac{1}{x}\right)'$$
$$= 2x\sin\frac{1}{x} - \cos\frac{1}{x}.$$

We cannot use this expression for $x = 0$. At the point $x = 0$ we can calculate the derivative using the definition of the derivative:

$$f'(0) = \lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x} = \lim_{\Delta x\to 0}\Delta x\sin\frac{1}{\Delta x} = 0$$

(the product of an infinitesimal function by a bounded function is an infinitesimal). Thus

$$f'(x) = \begin{cases} 2x\sin(1/x) - \cos(1/x) & \text{for } x \neq 0, \\ 0 & \text{for } x = 0. \end{cases}$$

### 1.5.9. The Logarithmic Derivative. The Derivative of a Power Function with Any Real Exponent. Table of Derivatives of the Basic Elementary Functions

**1. The concept of the logarithmic derivative of a function.** Let us calculate the derivative of the function $y = \ln |x|$ $(x \neq 0)$. Since $(\ln x)' = 1/x$ and $(\ln(-x))' = (-x)'/-x = 1/x$ (we have obtained the last equality on the basis of the rule of differentiation of a composite function), the derivative of the given function is expressed by the following formula:

$$y' = (\ln |x|)' = \frac{1}{x}. \tag{1}$$

Taking formula (1) into account, we shall calculate the derivative of the composite function $y = \ln |u|$, where $u = f(x)$ is a differentiable function. We have

$$y' = (\ln |u|)' = \frac{1}{u} u' = \frac{f'(x)}{f(x)}$$

or

$$(\ln |f(x)|)' = \frac{f'(x)}{f(x)}. \tag{2}$$

The derivative $(\ln |f(x)|)'$ is a *logarithmic derivative* of the function $f(x)$. To simplify the notation in logarithmic differentiation, the sign of the absolute value of the function $f(x)$ is usually omitted.

Using the logarithmic derivative, we shall calculate the derivative of the exponential power function $y = u(x)^{v(x)}$, where $u$ and $v$ are some functions of $x$ $(u > 0)$, which have derivatives $u'(x)$ and $v'(x)$ at the point $x$. Since $\ln y = v(x) \ln u(x)$, we can use formula (2) and obtain

$$\frac{y'}{y} = [v(x) \ln u(x)]' = v'(x) \ln u(x) + v(x) \frac{u'(x)}{u(x)}.$$

From this, taking into account that $y = u(x)^{v(x)}$, we get the following formula for the derivative of the exponential power function:

$$y' = u(x)^{v(x)} \left[ v'(x) \ln u(x) + v(x) \frac{u'(x)}{u(x)} \right]. \tag{3}$$

**Example.** Calculate the derivative of the function $y = x^x$.

*Solution.* This function can be represented as $y = u(x)^{v(x)}$, where $u(x) = x$ and $v(x) = x$. Using formula (3), we obtain

$$y' = x^x \left( 1 \ln x + x \frac{1}{x} \right) = x^x (\ln x + 1).$$

We can use another technique to calculate the derivative of the exponential power function $y = u(x)^{v(x)}$. We represent the function in the form $y = e^{v(x)\ln u(x)}$ and calculate $y'$:

$$y' = [e^{v(x)\ln u(x)}]' = e^{v(x)\ln u(x)}[v(x)\ln u(x)]'$$
$$= y\left[v'(x)\ln u(x) + v(x)\frac{u'(x)}{u(x)}\right],$$

substituting $y = u(x)^{v(x)}$, we again arrive at formula (3).

The logarithmic derivative is very convenient in seeking the derivative of a power function with any real exponent.

**2. The derivative of a power function with any real exponent.** The derivative of the function $y = x^\alpha$ ($\alpha$ is any real number) is expressed by the formula

$$y' = \alpha x^{\alpha-1}.$$

*Proof.* Since $y = x^\alpha$, it follows that $\ln y = \alpha \ln x$. From formula (2) we find that

$$\frac{y'}{y} = (\alpha \ln x)' = \frac{\alpha}{x}.$$

From this, taking into account that $y = x^\alpha$, we get a formula for the derivative of the power function:

$$y' = (x^\alpha)' = \alpha x^{\alpha-1}. \blacksquare$$

We have thus calculated the derivatives of all basic elementary functions and can compile the following table.

**3. Table of the derivatives of the basic elementary functions.**

I. $(C)' = 0$.

II. $(x^\alpha)' = \alpha x^{\alpha-1}$, in particular, $\left(\frac{1}{x}\right)' = -\frac{1}{x^2}$, $(\sqrt{x})' = \frac{1}{2\sqrt{x}}$.

III. $(\log_a x)' = \frac{1}{x}\log_a e$, in particular, $(\ln x)' = \frac{1}{x}$.

IV. $(a^x)' = a^x \ln a$, in particular, $(e^x)' = e^x$.

V. $(\sin x)' = \cos x$.

VI. $(\cos x)' = -\sin x$.

VII. $(\tan x)' = \frac{1}{\cos^2 x}$.

VIII. $(\cot x)' = -\frac{1}{\sin^2 x}$.

IX. $(\arcsin x)' = \frac{1}{\sqrt{1-x^2}}$.

X. $(\arccos x)' = -\frac{1}{\sqrt{1-x^2}}$.

XI. $(\arctan x)' = \frac{1}{1+x^2}$.

XII. $(\operatorname{arccot} x)' = -\frac{1}{1+x^2}$.

The formulas presented in the table as well as the rules of differentiation of the sum, the difference, the product and the quotient and the rule of differentiation of a composite function are the fundamental formulas of differential calculus. On the basis of the rules and formulas of differentiation we can make the following important conclusion: the derivative of any elementary function is also an elementary function. Thus the operation of differentiation does not take us outside the class of elementary functions.

### 1.5.10. Higher-Order Derivatives and Differentials

**1. The concept of a derivative of order $n$.** As was pointed out in 1.5.1, the derivative $f'(x)$ of the function $y = f(x)$ is itself a function of the argument $x$. Consequently, with respect to this derivative, we can again raise the question concerning the existence and the finding of the derivative.

We call $f'(x)$ the *first-order derivative* of the function $f(x)$.

The derivative of the derivative of some function is the *second-order derivative* (or the *second derivative*) of that function. The derivative of the second derivative is the *third-order derivative* (or the *third derivative*) and so on. Beginning with the second derivative, the derivatives are called *derivatives of higher orders* and are designated as $y''$, $y'''$, $y^{(4)}$, $y^{(5)}$, ..., $y^{(n)}$, ... (instead of $y''$ and $y'''$ we sometimes write $y^{(2)}$ and $y^{(3)}$), or $f''(x)$, $f'''(x)$, $f^{(4)}(x)$, $f^{(5)}(x)$, ..., $f^{(n)}(x)$, ....

The derivative of order $n$ (or the $n$th derivative) is the derivative of the derivative of order $(n-1)$, i.e. $y^{(n)} = (y^{(n-1)})'$.

Derivatives of higher orders have a variety of applications in physics. We shall give here the physical interpretation of the second derivative $f''(x)$. If the function $y = f(x)$ describes the law of motion of a particle along a straight line, then the first derivative $f'(x)$ is the instantaneous velocity of the particle at the moment $x$, and the second derivative is equal to the rate of change of the velocity, i.e. to the acceleration of the moving particle at this moment.

**2. Formulas for the $n$th derivatives of some functions.**

(1) Let us calculate the $n$th derivative of the power function $y = x^\alpha$ $(x > 0)$ ($\alpha$ is any real number). As a result of successive differentiation we have*

$$y' = \alpha x^{\alpha-1},\ y^{(2)} = \alpha(\alpha-1)\,x^{\alpha-2},\ y^{(3)} = \alpha(\alpha-1)(\alpha-2)\,x^{\alpha-3},\ \ldots,$$

$$y^{(n)} = \alpha(\alpha-1)(\alpha-2)\ \ldots\ [\alpha-(n-1)]\,x^{\alpha-n}.$$

* The strict derivation of the formulas of the $n$th derivatives must involve induction.

In a special case, when $\alpha = m$, where $m$ is a natural number, we obtain

$$(x^m)^{(m)} = m(m-1)(m-2)\ldots[m-(m-1)]\cdot 1 = m!,$$
$$(x^m)^{(n)} = 0 \text{ for } n > m.$$

(2) Let us calculate the $n$th derivative of the exponential function $y = a^x$ $(0 < a \neq 1)$. A successive differentiation results in

$$y' = a^x \ln a,\ y^{(2)} = a^x (\ln a)^2,\ y^{(3)} = a^x (\ln a)^3,\ \ldots,\ y^{(n)} = a^x (\ln a)^n.$$

In particular, if $y = e^x$, then we have, for any $n$,

$$(e^x)^{(n)} = e^x.$$

(3) Let us calculate now the $n$th derivative of the function $y = \sin x$. A successive differentiation results in

$$y' = \cos x = \sin\left(x + \frac{\pi}{2}\right),$$
$$y^{(2)} = -\sin x = \sin(x+\pi) = \sin\left(x + 2\frac{\pi}{2}\right),$$
$$y^{(3)} = -\cos x = \sin\left(x + 3\frac{\pi}{2}\right),\ \ldots,\ y^{(n)} = \sin\left(x + n\frac{\pi}{2}\right).$$

Thus we can calculate the derivative of any order of $\sin x$ by the formula

$$(\sin x)^{(n)} = \sin\left(x + n\frac{\pi}{2}\right).$$

For instance, $(\sin x)^{(10)} = \sin(x + 10(\pi/2)) = \sin(x+\pi) = -\sin x$.

(4) By analogy we can obtain a formula for the $n$th derivative of the function $y = \cos x$:

$$(\cos x)^{(n)} = \cos\left(x + n\frac{\pi}{2}\right).$$

**3. Leibniz' formula for finding the $n$th derivative of the product of two functions.** Let $y = u \cdot v$, where $u$ and $v$ are some functions of the variable $x$ which have derivatives of any order. Then

$$y' = u'v + uv',\ y'' = u''v + u'v' + u'v' + uv''$$
$$= u''v + 2u'v' + uv'' = u^{(2)}v + 2u'v' + uv^{(2)},$$
$$y''' = u'''v + u''v' + 2u''v' + 2u'v'' + u'v'' + uv'''$$
$$= u'''v + 3u''v' + 3u'v'' + uv''' = u^{(3)}v + 3u^{(2)}v' + 3u'v^{(2)} + uv^{(3)}.$$

The right-hand sides of the equations obtained resemble the expansions of different degrees of the binomial $(u+v)^n$ but the ex-

ponents are replaced by numbers which define the order of the derivative, and, for the analogy with the binomial formula to be complete, the functions $u$ and $v$ themselves must be regarded as "derivatives of order zero", i.e. $u^{(0)}$ and $v^{(0)}$. Taking this into account, we shall write the $n$th derivative of the product of two functions, in the general form, as follows:

$$y^{(n)} = (uv)^{(n)} = u^{(n)}v + nu^{(n-1)}v' + \frac{n(n-1)}{2!}u^{(n-2)}v'' + \\ \ldots + \frac{n(n-1)\ldots(n-k+1)}{k!}u^{(n-k)}v^{(k)} + \ldots + uv^{(n)}. \qquad (1)$$

Formula (1) is known as *Leibniz'* formula*. We shall use induction to prove this formula.

For $n = 1$ this formula assumes the form $(uv)' = u'v + uv'$, and this coincides with the formula for differentiation of the product of two functions. For $n = 2$ and $n = 3$ it has also been verified. It is therefore sufficient to assume the validity of formula (1) for a certain $n$ and prove its validity for $n + 1$. Let us differentiate this formula, i.e. find $y^{(n+1)} = (y^{(n)})'$:

$$y^{(n+1)} = u^{(n+1)}v + u^{(n)}v' + n[u^{(n)}v' + u^{(n-1)}v''] \\ + \frac{n(n-1)}{2!}[u^{(n-1)}v'' + u^{(n-2)}v'''] + \\ \ldots + \frac{n(n-1)\ldots(n-k+1)}{k!}[u^{(n-k+1)}v^{(k)} + u^{(n-k)}v^{(k+1)}] + \\ \ldots + u'v^{(n)} + uv^{(n+1)}.$$

Removing the brackets and collecting terms, we obtain

$$y^{(n+1)} = u^{(n+1)}v + (n+1)u^{(n)}v' + \left(n + \frac{n(n-1)}{2!}\right)u^{(n-1)}v'' + \\ \ldots + \left[\frac{n(n-1)\ldots(n-k+2)}{(k-1)!} + \frac{n(n-1)\ldots(n-k+1)}{k!}\right]u^{(n-k+1)}v^{(k)} + \\ \ldots + (n+1)u'v^{(n)} + uv^{(n+1)}.$$

But we can represent the expression in brackets as follows:

$$\frac{n(n-1)\ldots(n-k+2)}{(k-1)!} + \frac{n(n-1)\ldots(n-k+1)}{k!} \\ = \frac{n(n-1)(n-2)\ldots(n-k+2)(n-k+1)(n-k)(n-k-1)(n-k-2)\ldots 1}{(k-1)!(n-k+1)(n-k)(n-k-2)\ldots 1} \\ + \frac{n(n-1)\ldots(n-k+1)(n-k)(n-k-1)(n-k-2)\ldots 1}{k!(n-k)(n-k-1)(n-k-2)\ldots 1}$$

* Leibniz, Gottfried Wilhelm von (1646-1716), a German philosopher, physicist and mathematician.

$$= \frac{n!}{(k-1)!\,(n-k+1)!} + \frac{n!}{k!\,(n-k)!} = \frac{n!}{(k-1)!\,(n-k)!}\left(\frac{1}{n-k+1} + \frac{1}{k}\right)$$

$$= \frac{n!}{(k-1)!\,(n-k)!}\,\frac{n+1}{k!\,(n-k+1)} = \frac{(n+1)!}{k!\,(n-k+1)!}$$

$$= \frac{(n+1)\,n(n-1)\,(n-2)\,\ldots\,(n-k+2)\,(n-k+1)}{k!\,(n-k+1)\,(n-k)}$$

$$\times \frac{(n-k)\,(n-k-1)\,(n-k-2)\,\ldots\,1}{(n-k-1)\,(n-k-2)\,\ldots\,1} = \frac{(n+1)\,n(n-1)\,(n-2)\,\ldots\,(n-k+2)}{k!}.$$

Therefore

$$y^{(n+1)} = u^{(n+1)}v + (n+1)\,u^{(n)}v' + \frac{(n+1)\,n}{2!}\,u^{(n-1)}v'' +$$

$$\ldots + \frac{(n+1)\,n(n-1)\,\ldots\,(n-k+2)}{k!}\,u^{(n-k+1)}v^{(k)} +$$

$$\ldots + (n+1)\,u'v^{(n)} + uv^{(n+1)}.$$

We have completed the proof of formula (1). ■

**Example 1.** Calculate the fifth derivative of the function $y = x^5e^x$.

*Solution.* Setting $u = x^5$ and $v = e^x$, we find that $u' = 5x^4$, $u'' = 20x^3$, $u''' = 60x^2$, $u^{(4)} = 120x$, $u^{(5)} = 120$; $v' = v'' = v''' = v^{(4)} = v^{(5)} = e^x$. Substituting these expressions into formula (1) for $n = 5$, we obtain

$$y^{(5)} = 120e^x + 5\cdot 120xe^x + \frac{5\cdot 4}{1\cdot 2}\,60x^2e^x + \frac{5\cdot 4\cdot 3}{1\cdot 2\cdot 3}\,20x^3e^x + 5\cdot 5x^4e^x + x^5e^x$$

$$= e^x(120 + 600x + 600x^2 + 200x^3 + 25x^4 + x^5).$$

**Example 2.** Calculate the $n$th derivative ($n \geqslant 2$) of the function $y = x^2 \cos x$.

*Solution.* Setting $u = \cos x$ and $v = x^2$, we find that

$$u^{(n)} = \cos\left(x + n\frac{\pi}{2}\right), \quad v' = 2x, \quad v'' = 2, \quad v''' = v^{(4)} = v^{(5)} = \ldots = 0.$$

Substituting this into formula (1), we obtain

$$y^{(n)} = \cos\left(x + n\frac{\pi}{2}\right)x^2 + 2n\cos\left[x + (n-1)\frac{\pi}{2}\right]x$$

$$+ \frac{2n(n-1)}{1\cdot 2}\cos\left[x + (n-2)\frac{\pi}{2}\right].$$

**4. Differentials of higher orders.** Let us consider differentials of higher orders. For the sake of convenience, we shall use the designations $\delta y$ and $\delta x$ alongside the designations $dy$ and $dx$.

Let the function $f(x)$ be differentiable at every point $x$ of a certain closed interval. Then its differential

$$dy = f'(x)\,dx,$$

which we shall call the *first-order differential*, is the function of two variables of the argument $x$ and of its differential $dx$. Let the function $f'(x)$, in its turn, be differentiable at a point $x$. We consider $dx$ in the expression for $dy$ to be a constant factor. Then the function $dy$ is the function of only the argument $x$ and its differential at the point $x$ has the form (when considering the differential of $dy$ we shall use the new designations of differentials)

$$\delta(dy) = \delta[f'(x)\,dx] = [f'(x)\,dx]'\,\delta x = f''(x)\,dx\,\delta x.$$

The differential $\delta(dy)$ of the differential $dy$ at the point $x$, taken for $\delta x = dx$, is the *second-order differential* of the function $f(x)$ at the point $x$ and is designated as $d^2y$, i.e.

$$d^2y = f''(x)\,(dx)^2.$$

In its turn, the differential $\delta(d^2y)$ of the differential $d^2y$, taken for $\delta x = dx$, is the *third-order differential* of the function $f(x)$ and is designated as $d^3y$, and so on. The differential $\delta(d^{n-1}y)$ of the differential $d^{n-1}y$, taken for $\delta x = dx$, is the *nth-order differential* (or the *nth differential*) of the function $f(x)$ and is designated as $d^ny$.

We shall prove that the formula

$$d^ny = y^{(n)}(x)\,(dx)^n,\ n = 1, 2, \ldots \tag{2}$$

is valid for the $n$th-order differential of the function. We shall carry out the proof by induction. For $n = 1$ and $n = 2$ formula (2) has been proved. Assume that it holds true for the differential of order $n - 1$:

$$d^{n-1}y = y^{(n-1)}(x)\,(dx)^{n-1},$$

and the function $y^{(n-1)}(x)$, in its turn, is differentiable at a point $x$. Then

$$d^ny = \delta(d^{n-1}y) = \delta[y^{(n-1)}(x)\,(dx)^{n-1}]$$
$$= [y^{(n-1)}(x)\,(dx)^{n-1}]'\,\delta x = y^{(n)}(x)\,\delta x\,(dx)^{n-1}.$$

Setting $\delta x = dx$, we obtain

$$d^ny = \delta(d^{n-1}y)\,|_{\delta x = dx} = y^{(n)}(x)\,(dx)^n,$$

and this is what we wished to prove.

It follows from formula (2) that for any $n$ there holds a formula

$$y^{(n)}(x) = \frac{d^ny}{(dx)^n},\ \text{or}\ y^{(n)}(x) = \frac{d^ny}{dx^n},$$

i.e. the $n$th derivative of the function $y = f(x)$ at the point $x$ is equal to the ratio of the $n$th differential of this function at the point $x$ to the $n$th degree of the differential of the argument.

**Example 3.** Calculate the differential $d^3y$ of the function $y = x^4 - 3x^2 + 4$.

*Solution.* A successive differentiation results in

$$y'(x) = 4x^3 - 6x,\ y''(x) = 12x^2 - 6,\ y'''(x) = 24x$$

and, consequently, $d^3y = y'''(x)(dx)^3 = 24x(dx)^3$.

### 1.5.11. The Parametric Representation of a Function and Its Differentiation

**1. The parametric representation of a function.** We are given two functions

$$x = \varphi(t) \text{ and } y = \psi(t) \tag{1}$$

of one independent variable $t$ defined and continuous in the same closed interval. If $x = \varphi(t)$ is strictly monotone, then its inverse $t = \Phi(x)$ is single-valued, also continuous and strictly monotone. We can therefore regard $y$ as a function which depends on the variable $x$ by means of the variable $t$, called a *parameter*:

$$y = \psi[\Phi(x)].$$

In this case we say that the function $y$ of $x$ is represented parametrically by means of equations (1).

Note that the function $\psi[\Phi(x)]$ is continuous by virtue of the theorem on the continuity of a composite function.

**Example 1.** Let $x = R\cos t$, $y = R\sin t$ $(0 \leqslant t \leqslant \pi)$. Since the function $x = R\cos t$ decreases for $0 \leqslant t \leqslant \pi$, these equations represent parametrically the function $y$ of $x$. If we express $t$ in terms of $x$ from the first equation and substitute the expression obtained into the second equation, we shall get the required function of the variable $x$ in explicit form.

We can do this even easier if we note that

$$x^2 + y^2 = R^2(\cos^2 t + \sin^2 t) = R^2.$$

Hence $y = \sqrt{R^2 - x^2}$ or $y = -\sqrt{R^2 - x^2}$. Since the function $y = R\sin t$ is nonnegative for $0 \leqslant t \leqslant \pi$, we choose the plus sign before the radical, i.e. $y = \sqrt{R^2 - x^2}$.

If $\pi \leqslant t \leqslant 2\pi$, then $y = -\sqrt{R^2 - x^2}$.

We can thus infer that when $t$ varies from 0 to $2\pi$, the formulas $x = R\cos t$ and $y = R\sin t$ define two functions of the variable $x$, whose graphs form a circle of radius $R$.

**Example 2.** Let $x = a\cos t$, $y = b\sin t$ $(0 \leqslant t \leqslant 2\pi)$.

These equations are the parametric equations of the ellipse since (see the remark in Item 1 of 1.3.7) the ellipse results when a circle of radius $a$ is contracted $a/b$ times along the $y$-axis. It follows from Example 1 that the equations $x = a\cos t$, $y = a\sin t$ $(0 \leqslant t \leqslant 2\pi)$ are the parametric equations of the circle $x^2 + y^2 = a^2$. Thus the parametric equations of the ellipse result from the parametric

equations of the circle when the right-hand side of the equation for the ordinate $y$ is multiplied by $b/a$ and have the form $x = a \cos t$, $y = b \sin t$ $(0 \leqslant t \leqslant 2\pi)$. The procedure can be even simpler. Excluding the parameter $t$ from these equations (solving them for $\cos t$ and $\sin t$, squaring the equations obtained and adding them together), we get

$$(x/a)^2 + (y/b)^2 = \cos^2 t + \sin^2 t = 1,$$

or

$$x^2/a^2 + y^2/b^2 = 1,$$

which is the equation of the ellipse.

The parametric representation of a function is especially important when we study the motion of a particle. If the particle moves on the plane, then its coordinates $x$ and $y$ are the functions of time $t$. Defining these functions $x = \varphi(t)$ and $y = \psi(t)$, we shall completely define the motion of the particle. For every time interval in which the function $\varphi(t)$ is strictly monotone, we can, as before, define the function $y = \psi[\Phi(x)]$, whose graph is a curve described by the moving particle in this time interval. In the last example the functions described the motion of the particle along the ellipse.

**2. Differentiation of a function represented parametrically.** Let us now assume that the functions $x = \varphi(t)$ and $y = \psi(t)$ have derivatives and $\varphi'(t) \neq 0$ on some interval. It follows from the last inequality (we shall show this) that the function $x = \varphi(t)$ is strictly monotone (see Theorem 6.7, Chapter 6) and, consequently, the inverse function $t = \Phi(x)$ is single-valued. According to Theorem 5.4 on the derivative of an inverse function, the function $\Phi(x)$ has a derivative

$$\Phi'(x) = \frac{1}{\varphi'(t)},$$

and according to Theorem 5.5 on the derivative of a composite function, the function $y = \psi[\Phi(x)]$ has a derivative

$$y'(x) = \psi'[\Phi(x)]\,\Phi'(x).$$

Consequently,

$$y'(x) = \frac{\psi'(t)}{\varphi'(t)}\bigg|_{t=\Phi(x)}. \tag{2}$$

We have thus proved that the derivative of a function represented parametrically is expressed by formula (2).

**Example 1.** Find $y'(x)$ if $x = R\cos t$, $y = R\sin t$ $(0 \leqslant t \leqslant \pi)$.

*Solution.* From formula (2) we get (here $t = \Phi(x) = \arccos(x/R)$)

$$y'(x) = \frac{R\cos t}{-R\sin t}\bigg|_{t=\arccos(x/R)} = -\frac{\cos t}{\sqrt{1-\cos^2 t}}\bigg|_{t=\arccos(x/R)}$$

$$= -\frac{x/R}{\sqrt{1-x^2/R^2}} = -\frac{x}{\sqrt{R^2-x^2}} \quad (x \neq \pm R).$$

If we use the explicit expression for the function $y$ of $x$: $y = \sqrt{R^2 - x^2}$, then we obviously obtain the same result:

$$y'(x) = \frac{-2x}{2\sqrt{R^2-x^2}} = -\frac{x}{\sqrt{R^2-x^2}} \quad (x \neq \pm R).$$

Assume that there are second derivatives of the functions $\varphi'(t)$ and $\psi'(t)$ at a point $t$. Then we can calculate the second derivative of the function represented parametrically. Note that the function $y'(x) = \frac{\psi'(t)}{\varphi'(t)}\Big|_{t=\Phi(x)}$, in its turn, is represented parametrically by the equation $y'_x = \frac{\psi'(t)}{\varphi'(t)} = \psi_1(t)$ and $x = \varphi(t)$. Consequently, from formula (2) we have

$$y''(x) = (y'(x))'_x = \frac{\psi'_1(t)}{\varphi'(t)}\Big|_{t=\Phi(x)} = \frac{\left(\frac{\psi'(t)}{\varphi'(t)}\right)'}{\varphi'(t)}\Bigg|_{t=\Phi(x)}$$

$$= \frac{\frac{\psi''(t)\,\varphi'(t) - \varphi''(t)\,\psi'(t)}{[\varphi'(t)]^2}}{\varphi'(t)}\Bigg|_{t=\Phi(x)} = \frac{\psi''(t)\,\varphi'(t) - \varphi''(t)\,\psi'(t)}{[\varphi'(t)]^3}\Bigg|_{t=\Phi(x)}.$$

We have used here the rule of differentiation of a quotient. Thus

$$y''(x) = \frac{\psi''(t)\,\varphi'(t) - \varphi''(t)\,\psi'(t)}{[\varphi'(t)]^3}\Bigg|_{t=\Phi(x)}. \tag{3}$$

By analogy we can get the derivative of $y$ with respect to $x$ of any order.

**Example 2.** Find $y''(x)$ if $x = \cos t$, $y = \sin t$ $(0 \leqslant t \leqslant \pi)$.

*Solution.* $y'(t) = \cos t$, $y''(t) = -\sin t$, $x'(t) = -\sin t$, $x''(t) = -\cos t$. Substituting these expressions into formula (3), we find that

$$y''(x) = \frac{(-\sin t)(-\sin t) - (-\cos t)(\cos t)}{(-\sin t)^3}\Bigg|_{t=\arccos(x/R)}$$

$$= \frac{\sin^2 t + \cos^2 t}{(-\sin t)^3}\Bigg|_{t=\arccos(x/R)} = -\frac{1}{\sin^3 t}\Bigg|_{t=\arccos(x/R)} = -\frac{1}{(1-x^2/R^2)^{3/2}}.$$

## Chapter 6

## INVESTIGATING FUNCTIONS BY DIFFERENTIATION

In the preceding chapter we got acquainted with differentiation of functions. Let us consider now the methods of investigating the behaviour of functions and constructing graphs which are widely used both in the theory and in practical application.

### 1.6.1. Fundamental Theorems of Differential Calculus

**Theorem 6.1.** *Assume that the function $f(x)$ is defined on the open interval $(a, b)$ and has the greatest or the least value at a point $x_0$ of this interval. Then, if there is a derivative at the point $x_0$, then it is equal to zero, i.e. $f'(x_0) = 0$.*

*Proof.* Assume, for definiteness, that the function $f(x)$ has the greatest value at the point $x_0$, i.e. $f(x) \leqslant f(x_0)$ for any $x \in (a, b)$.

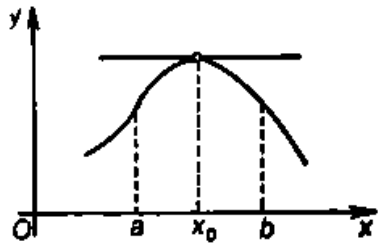



Fig. 69

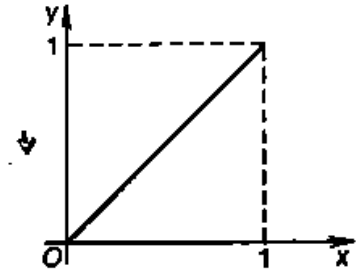



Fig. 70

This means that $\Delta y = f(x_0 + \Delta x) - f(x_0) \leqslant 0$ for any point $x_0 + \Delta x \in (a, b)$. Therefore, if $\Delta x > 0$ $(x > x_0)$, then $\Delta y/\Delta x \leqslant 0$, and, consequently,

$$f'_+(x_0) = \lim_{\Delta x \to 0+} \frac{\Delta y}{\Delta x} \leqslant 0,$$

but if $\Delta x < 0$ $(x < x_0)$, then $\Delta y/\Delta x \geqslant 0$, and, consequently,

$$f'_-(x_0) = \lim_{\Delta x \to 0-} \frac{\Delta y}{\Delta x} \geqslant 0,$$

i.e. the right-hand derivative at the point $x_0$ is nonpositive and the left-hand derivative is nonnegative. By the hypothesis, $f'(x_0)$ exists and, hence, $f'_+(x_0) = f'_-(x_0) = f'(x_0)$. This is possible only if $f'_+(x_0) = f'_-(x_0) = 0$. But then $f'(x_0) = 0$ as well.

The case when the function $f(x)$ has the least value at the point $x_0$ can be treated by analogy. ■

The geometric meaning of Theorem 6.1 is that if the differentiable function $f(x)$ has the greatest or the least value at the point $x_0$, then at the point $(x_0, f(x_0))$ the tangent to the graph of the function $f(x)$ is parallel to the $x$-axis (Fig. 69).

**Remark.** The theorem is not true if the function $f(x)$ is considered on the closed interval $[a, b]$. Thus, for instance, on the closed interval $[0, 1]$ the function $f(x) = x$ assumes the least value at the point $x = 0$ and the greatest value at the point $x = 1$. However, at either point the derivative does not vanish but is equal to unity (Fig. 70).

**Theorem 6.2 (Rolle's* theorem).** *Assume that the function $f(x)$ is defined on $[a, b]$ and* (1°) *$f(x)$ is continuous on $[a, b]$;* (2°) *$f(x)$ is differentiable on $(a, b)$;* (3°) *$f(a) = f(b)$. Then there is a point $c \in (a, b)$ at which $f'(c) = 0$.*

*Proof.* Since the function $f(x)$ is continuous on $[a, b]$, it follows, according to Weierstrass's second theorem, that it has the maximum value $M$ and the minimum value $m$ on this interval, i.e. there are points $x_1, x_2 \in [a, b]$ such that $f(x_1) = m$, $f(x_2) = M$, and the inequalities

$$m \leqslant f(x) \leqslant M$$

are satisfied.

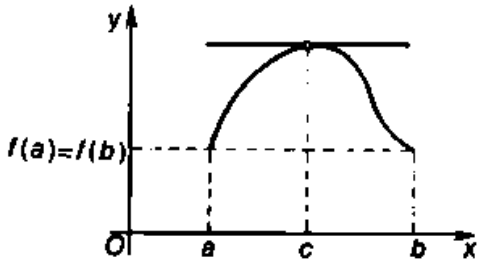


Fig. 71

Two cases are possible here: (1) $M = m$; (2) $m < M$.

In the first case $f(x) = \text{const} = M = m$. Therefore the derivative $f'(x)$ is zero at any point of $[a, b]$ and the theorem is proved.

In the second case, since $f(a) = f(b)$, at least one of the two values, $m$ or $M$, is not assumed by the function at the endpoints of the interval $[a, b]$, i.e. there is a point $c \in (a, b)$ at which the function $f(x)$ assumes the greatest or the least value on the interval $(a, b)$. Then, since $f(x)$ is differentiable at the point $c$, it follows from Theorem 6.1 that $f'(c) = 0$. ■

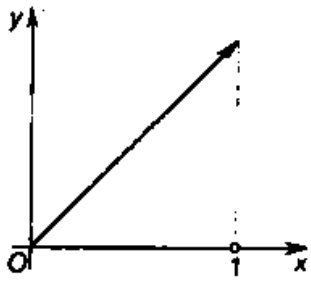


Fig. 72

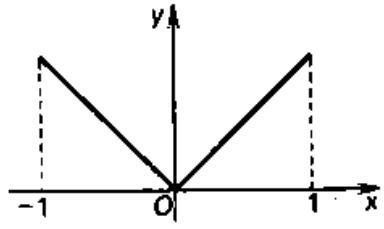


Fig. 73

From the point of view of geometry, Rolle's theorem means that the graph of a function, which is continuous on the interval $[a, b]$ and differentiable within this interval and which assumes equal values at its endpoints, includes a point $(c, f(c))$ at which the tangent is parallel to the $x$-axis (Fig. 71). In Fig. 71 the function $f(x)$ assumes the greatest value at the point $c$.

It should be pointed out that all the three conditions of Rolle's theorem are essential. To ascertain this, it is sufficient to consider functions for which two conditions of the theorem would be fulfilled and the third condition would not be fulfilled and whose de-

* Rolle, Michel (1652-1719), a French mathematician.

rivatives would not vanish at any point. Thus, for instance, the function $f(x) = x$, $x \in [0, 1]$ (see Fig. 70), satisfies conditions 1° and 2° but does not satisfy condition 3° and there is no point $c$ for it such that $f'(c) = 0$. Let us consider two more examples.

The function $f(x)$, equal to $x$ if $0 \leqslant x < 1$ and to 0 if $x = 1$ (Fig. 72), satisfies conditions 2° and 3° but does not satisfy condition 1°. The function $f(x) = |x|$, $x \in [-1, 1]$ (Fig. 73), satisfies conditions 1° and 3° but does not satisfy condition 2°. There is no point, for these functions, at which their derivative would vanish.

Note that in mathematics the fact that some particular conditions of the theorems being proved are essential is verified by constructing requisite examples which show that when this or that condition of the theorem is not fulfilled, the statement of the theorem becomes untrue.

**Theorem 6.3 (Lagrange's* theorem).** *Assume that the function $f(x)$ is defined on $[a, b]$, and (1°) $f(x)$ is continuous on $[a, b]$; (2°) $f(x)$ is differentiable on $(a, b)$. Then there is a point $c \in (a, b)$ such that the formula*

$$\frac{f(b)-f(a)}{b-a} = f'(c)$$

*holds true.*

*Proof.* We introduce into consideration, on $[a, b]$, an auxiliary function

$$F(x) = f(x) - f(a) - \frac{f(b)-f(a)}{b-a}(x-a).$$

The function $F(x)$ satisfies all the three conditions of Rolle's theorem:

(1) $F(x)$ is continuous on $[a, b]$ $\Big($being the difference of two continuous functions $f(x)$ and the linear function $f(a) + \frac{f(b)-f(a)}{b-a} \times (x-a)\Big)$;

(2) $F(x)$ is differentiable on $(a, b)$, i.e. inside $[a, b]$ it has a derivative equal to $F'(x) = f'(x) - \frac{f(b)-f(a)}{b-a}$;

(3) $F(a) = 0$ and $F(b) = 0$, i.e. $F(a) = F(b)$.

Consequently, in accordance with Rolle's theorem, there is a point $c \in (a, b)$ such that $F'(c) = 0$, i.e. $f'(c) - \frac{f(b)-f(a)}{b-a} = 0$. From this we find that $f'(c) = \frac{f(b)-f(a)}{b-a}$. ■

* Lagrange, Joseph Louis (1736-1813), a French mathematician, physicist and astronomer.

We shall establish the geometric meaning of Lagrange's theorem (Fig. 74). The quantity $\frac{f(b)-f(a)}{b-a}$ is the slope of the secant which passes through the points $M_1(a, f(a))$ and $M_2(b, f(b))$ of the graph of the function $y = f(x)$, and $f'(c)$ is the slope of the tangent to the graph at the point $(c, f(c))$. It follows from Lagrange's theorem that there is a point $c$ such that the tangent to the graph at the point $(c, f(c))$ is parallel to the secant $M_1M_2$. There may be several points of this kind, but at least one always exists.

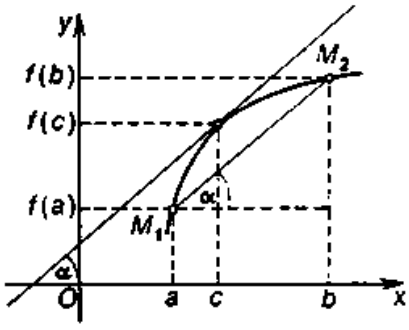



Fig. 74

**Remark 1.** The relation

$$f(b) - f(a) = f'(c)(b-a), \quad a < c < b, \tag{1}$$

is known as *Lagrange's formula*, or the *mean value theorem (law of the mean)*.

**Remark 2.** Since the point $c$ lies between $a$ and $b$, it follows that $c = a + \theta(b-a)$, where $0 < \theta < 1$. Taking this formula into account, we can write Lagrange's formula as

$$f(b) - f(a) = f'(a + \theta(b-a))(b-a), \text{ where } 0 < \theta < 1.$$

**Remark 3.** If we set $a = x$ and $b = x + \Delta x$, we get

$$f(x + \Delta x) - f(x) = f'(x + \theta\Delta x)\Delta x, \text{ where } 0 < \theta < 1.$$

This notation for Lagrange's formula is often more convenient than notation (1).

We shall show later that Lagrange's theorem is at the basis of the proof of many formulas and theorems of analysis.

**Theorem 6.4 (Cauchy's theorem).** *Assume that the functions $f(x)$ and $g(x)$ are continuous on $[a, b]$ and differentiable on $(a, b)$. Assume, in addition, that $g'(x) \neq 0$. Then there is a point $c \in (a, b)$ such that the formula*

$$\frac{f(b)-f(a)}{g(b)-g(a)} = \frac{f'(c)}{g'(c)} \tag{2}$$

*holds true.*

*Proof.* We shall first show that $g(b) \neq g(a)$, i.e. that formula (2) has sense. Indeed, if we assume that $g(b) = g(a)$, then, in accordance with Rolle's theorem, for the function $g(x)$ we can find a point $\xi \in (a, b)$ at which $g'(\xi) = 0$. And this contradicts the condition $g'(x) \neq 0$ on $(a, b)$. Let us prove formula (2).

We shall consider, on $[a, b]$, an auxiliary function

$$F(x) = f(x) - f(a) - \frac{f(b)-f(a)}{g(b)-g(a)}[g(x) - g(a)].$$

It is easy to notice that on $[a, b]$ the function $F(x)$ satisfies the conditions of Rolle's theorem. Indeed, $F(x)$ is continuous on $[a, b]$, differentiable on $(a, b)$, and, in addition, the substitution $x = a$ and $x = b$ yields $F(a) = 0$ and $F(b) = 0$, i.e. $F(a) = F(b)$. According to Rolle's theorem, for $F(x)$ there is a point $c$, $a < c < b$, such that $F'(c) = 0$.

Since $F'(x) = f'(x) - \frac{f(b)-f(a)}{g(b)-g(a)} g'(x)$, it follows that

$$F'(c) = f'(c) - \frac{f(b)-f(a)}{g(b)-g(a)} g'(c) = 0.$$

Whence, taking into account that $g'(c) \neq 0$, we get formula (2). ■

**Remark.** Formula (2) is known as *Cauchy's formula*, or *generalized mean value theorem*.

### 1.6.2. Evaluating Indeterminate Forms. L'Hôpital's Rule

**1. Evaluating the indeterminate form $\frac{0}{0}$.** We say that the ratio of two functions $\frac{f(x)}{g(x)}$, as $x \to a$, is an indeterminate form $\frac{0}{0}$ if

$$\lim_{x\to a} f(x) = \lim_{x\to a} g(x) = 0.$$

To evaluate this indeterminate form is to calculate the limit $\lim_{x\to a} \frac{f(x)}{g(x)}$, provided that it exists, or to establish its nonexistence. The following theorem establishes a rule for evaluating the indeterminate form $\frac{0}{0}$.

**Theorem 6.5 (L'Hôpital's* rule).** *Assume that the functions $f(x)$ and $g(x)$ are defined and differentiable in some neighbourhood of the point $a$, except maybe for the point $a$ itself. Assume furthermore that $\lim_{x\to a} f(x) = \lim_{x\to a} g(x) = 0$** and $g'(x) \neq 0$ in the indicated neighbourhood of the point $a$. Then, if there is a limit of the ratio of the derivatives $\lim_{x\to a} \frac{f'(x)}{g'(x)}$ (finite or infinite), then there also exists a limit $\lim_{x\to a} \frac{f(x)}{g(x)}$, and the formula*

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}$$

*holds true.*

* L'Hôpital, Guillaume François Antoine de (1661-1704), a French mathematician.

** The theorem remains valid in the case when $x \to a-$ or $x \to a+$.

*Proof.* Assume that $\{x_n\}$ is an arbitrary sequence of the values of the argument, which converges to the point $a$, and $x_n \neq a$. We complete the definition of the functions $f(x)$ and $g(x)$ at the point $a$ setting them equal to zero, i.e. $f(a) = g(a) = 0$. Then, evidently, the functions $f(x)$ and $g(x)$ are continuous on $[a, x_n]$, differentiable on $(a, x_n)$ and, by the hypothesis, $g'(x) \neq 0$. Thus, for $f(x)$ and

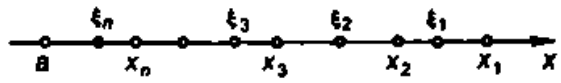



Fig. 75

$g(x)$, all the conditions of Cauchy's theorem are fulfilled on $[a, x_n]$, i.e. there is a point $\xi_n$ within $[a, x_n]$ such that

$$\frac{f(x_n)-f(a)}{g(x_n)-g(a)} = \frac{f'(\xi_n)}{g'(\xi_n)}, \quad \xi_n \in (a, x_n).$$

When we complete the definition, we get $f(a) = g(a) = 0$, and, consequently,

$$\frac{f(x_n)}{g(x_n)} = \frac{f'(\xi_n)}{g'(\xi_n)}, \quad \xi_n \in (a, x_n). \tag{1}$$

Let now $n \to \infty$ in formula (1). Then, evidently, $\xi_n \to a$ as $n \to \infty$ (Fig. 75). Since $\lim\limits_{x\to a} \frac{f'(x)}{g'(x)}$ exists, the right-hand side of formula (1) has a limit equal to $\lim\limits_{x\to a} \frac{f'(x)}{g'(x)}$ as $n \to \infty$. Consequently, as $n \to \infty$, there is a limit of the left-hand side of formula (1) as well, and

$$\lim_{n\to\infty} \frac{f(x_n)}{g(x_n)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}.$$

Since $\{x_n\}$ is an arbitrary sequence of the values of the argument, which converges to $a$, we can infer that $\lim\limits_{x\to a} \frac{f(x)}{g(x)}$ exists and $\lim\limits_{x\to a} \frac{f(x)}{g(x)} = \lim\limits_{x\to a} \frac{f'(x)}{g'(x)}$. ■

This theorem is known as *L'Hôpital's rule.*

**Remark 1.** If the derivatives $f'(x)$ and $g'(x)$ satisfy the same requirements as the functions $f(x)$ and $g(x)$ themselves, we can apply L'Hôpital's rule repeatedly, and then we obtain

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)} = \lim_{x\to a} \frac{f''(x)}{g''(x)}.$$

**Remark 2.** The theorem remains valid in the case when $x \to \infty$, $x \to +\infty$, $x \to -\infty$. Indeed, assume, for instance, that $\lim\limits_{x\to\infty} f(x) =$

$\lim\limits_{x\to\infty} g(x)=0$ and $\lim\limits_{x\to\infty}\frac{f'(x)}{g'(x)}$ exists (finite or infinite). We make a substitution $x = 1/t$, and then $t \to 0$ as $x \to \infty$ and $f(x) = f(1/t) \to 0$, $g(x) = g(1/t) \to 0$ as $t \to 0$. Applying Theorem 6.5 and the rule of differentiation of a composite function to the functions $f(1/t)$ and $g(1/t)$, we obtain

$$\lim_{x\to\infty}\frac{f(x)}{g(x)} = \lim_{t\to 0}\frac{f(1/t)}{g(1/t)} = \lim_{t\to 0}\frac{f'(1/t)(-1/t^2)}{g'(1/t)(-1/t^2)}$$
$$= \lim_{t\to 0}\frac{f'(1/t)}{g'(1/t)} = \lim_{x\to\infty}\frac{f'(x)}{g'(x)}.$$

Here are some examples.

1. $\lim\limits_{x\to 0}\frac{1-\cos x}{x^2} = \lim\limits_{x\to 0}\frac{\sin x}{2x} = \frac{1}{2}\lim\limits_{x\to 0}\frac{\sin x}{x} = \frac{1}{2}\cdot 1 = \frac{1}{2}$.

2. $\lim\limits_{x\to 0}\frac{x-\sin x}{x^3} = \lim\limits_{x\to 0}\frac{1-\cos x}{3x^2} = \lim\limits_{x\to 0}\frac{\sin x}{6x} = \frac{1}{6}\lim\limits_{x\to 0}\frac{\sin x}{x}$
$= \frac{1}{6}\cdot 1 = \frac{1}{6}$.

3. $\lim\limits_{x\to 0}\frac{e^x-1}{x} = \lim\limits_{x\to 0}\frac{e^x}{1} = \lim\limits_{x\to 0} e^x = 1$.

**2. Evaluating the indeterminate form $\frac{\infty}{\infty}$.** We say that the ratio of two functions $\frac{f(x)}{g(x)}$, as $x \to a$, is an indeterminate form $\frac{\infty}{\infty}$ if

$$\lim_{x\to a} f(x) = \lim_{x\to a} g(x) = \infty, \ +\infty \text{ or } -\infty.$$

A statement similar to Theorem 6.5 is true for this indeterminate form, namely, if, in the formulation of the theorem, we replace the requirement $\lim\limits_{x\to a} f(x) = \lim\limits_{x\to a} g(x) = 0$ by the requirement $\lim\limits_{x\to a} f(x) = \lim\limits_{x\to a} g(x) = \infty$, then the theorem remains true.

Here are some examples.

1. $\lim\limits_{x\to+\infty}\frac{\ln x}{x} = \lim\limits_{x\to+\infty}\frac{(1/x)}{1} = \lim\limits_{x\to+\infty}\frac{1}{x} = 0$.

2. $\lim\limits_{x\to+\infty}\frac{x^n}{e^x} = \lim\limits_{x\to+\infty}\frac{nx^{n-1}}{e^x} = \lim\limits_{x\to+\infty}\frac{n(n-1)x^{n-2}}{e^x} = \ldots = \lim\limits_{x\to\infty}\frac{n!}{e^x} = 0$.

3. $\lim\limits_{x\to 0+}\sqrt{x}\ln x = \lim\limits_{x\to 0+}\frac{\ln x}{x^{-1/2}} = \lim\limits_{x\to 0+}\frac{1/x}{(-1/2)x^{-3/2}}$
$= -2\lim\limits_{x\to 0+}\sqrt{x} = 0$.

**3. Other indeterminate forms and their evaluation.** The indeterminate forms $0\cdot\infty$ and $\infty - \infty$ can be reduced to the indeterminate forms $\frac{0}{0}$ and $\frac{\infty}{\infty}$. The following examples show this.

**Example 1.** Find $\lim\limits_{x\to 0+} x\ln x$.

*Solution.* We have an indeterminate form $0\cdot\infty$. But $x\ln x = \ln x/(1/x)$, and we have got an indeterminate form $\frac{\infty}{\infty}$. Applying L'Hôpital's rule, we obtain

$$\lim_{x\to 0+} x\ln x = \lim_{x\to 0+}\frac{(\ln x)'}{(1/x)'} = \lim_{x\to 0+}\frac{1/x}{-1/x^2} = -\lim_{x\to 0+} x = 0.$$

**Example 2.** Find $\lim\limits_{x\to\pi/2}(\sec x - \tan x)$.

*Solution.* We have an indeterminate form $\infty-\infty$. But $\sec x - \tan x = \frac{1}{\cos x} - \frac{\sin x}{\cos x} = \frac{1-\sin x}{\cos x}$, and we have got an indeterminate form $\frac{0}{0}$ under the same condition $x\to\pi/2$. Applying L'Hôpital's rule, we obtain

$$\lim_{x\to\pi/2}(\sec x - \tan x) = \lim_{x\to\pi/2}\left(\frac{1-\sin x}{\cos x}\right) = \lim_{x\to\pi/2}\frac{-\cos x}{-\sin x} = 0.$$

Let us consider now the indeterminate forms $0^0$, $1^\infty$, $\infty^0$. We come across indeterminate forms of these types when we consider functions of the form $y = f(x)^{g(x)}$ if, as $x\to a$, the function $f(x)$ tends to 0, 1 and $\infty$ respectively, and $g(x)$ tends to 0, $\infty$ and 0 respectively. Using the identity

$$f(x)^{g(x)} = e^{g(x)\ln f(x)},$$

we can reduce these indeterminate forms to the indeterminate form $0\cdot\infty$, which we have considered earlier.

**Example 3.** Find $\lim\limits_{x\to 0+} x^x$.

*Solution.* We have an indeterminate form $0^0$. But $x^x = e^{x\ln x}$, and we have got an indeterminate form $0\cdot\infty$ in the exponent which we have already considered (see Example 1). Consequently,

$$\lim_{x\to 0+} x^x = \lim_{x\to 0+} e^{x\ln x} = e^{\lim\limits_{x\to 0+} x\ln x} = e^0 = 1.$$

**Example 4.** Find $\lim\limits_{x\to 0}(1+x^2)^{\frac{1}{e^x-1-x}}$.

*Solution.* We have an indeterminate form $1^\infty$. But

$$(1+x^2)^{\frac{1}{e^x-1-x}} = e^{\frac{\ln(1+x^2)}{e^x-1-x}},$$

and we have got an indeterminate form $\frac{0}{0}$ in the exponent.

Applying L'Hôpital's rule, we obtain

$$\lim_{x\to 0}\frac{\ln(1+x^2)}{e^x-1-x}=\lim_{x\to 0}\frac{2x/(1+x^2)}{e^x-1}$$

$$=\lim_{x\to 0}\frac{2x}{(e^x-1)(1+x^2)}=\lim_{x\to 0}\frac{2}{e^x(1+x^2)+(e^x-1)2x}=\frac{2}{1}=2.$$

Consequently,

$$\lim_{x\to 0}(1+x^2)^{\frac{1}{e^x-1-x}}=e^{\lim\limits_{x\to 0}\frac{\ln(1+x^2)}{e^x-1-x}}=e^2.$$

**Example 5.** Find $\lim\limits_{x\to\pi/2}(\tan x)^{2\cos x}$.

*Solution.* We have an indeterminate form $\infty^0$. But

$$(\tan x)^{2\cos x}=e^{2\cos x\ln\tan x}=e^{\frac{2\ln\tan x}{1/(\cos x)}},$$

and we have got an indeterminate form $\frac{\infty}{\infty}$ in the exponent. Applying L'Hôpital's rule, we obtain

$$\lim_{x\to\pi/2}\frac{2\ln\tan x}{1/(\cos x)}=2\lim_{x\to\pi/2}\frac{\ln\tan x}{\sec x}=2\lim_{x\to\pi/2}\frac{1/(\tan x)\sec^2 x}{\sec x\tan x}$$

$$=2\lim_{x\to\pi/2}\frac{\sec x}{\tan^2 x}=2\lim_{x\to\pi/2}\frac{\sec x\tan x}{2\tan x\sec^2 x}=\lim_{x\to\pi/2}\cos x=0.$$

Consequently,

$$\lim_{x\to\pi/2}(\tan x)^{2\cos x}=e^{\lim\limits_{x\to\pi/2}2\cos x\ln\tan x}=e^0=1.$$

### 1.6.3. Taylor's Formula

We shall consider now one of the fundamental formulas of analysis which has a variety of applications in the analysis itself and in related fields of science.

**1. Taylor's formula.**

**Theorem 6.6 (Taylor's* theorem).** *Assume that the function $f(x)$ has derivatives of order $n+1$ at the point $a$ and in its certain neighbourhood.** Let $x$ be any value of the argument belonging to the indicated neighbourhood, $x\neq a$. Then between the points $a$ and $x$ there*

* Taylor, Brook (1685-1731), an English mathematician.

** From this it follows that the function $f(x)$ itself and its derivatives up to the order $n$ inclusive are continuous and differentiable in this neighbourhood.

*is a point $\xi$ such that the following formula*

$$f(x)=f(a)+\frac{f'(a)}{1!}(x-a)+\frac{f''(a)}{2!}(x-a)^2+$$
$$\ldots+\frac{f^{(n)}(a)}{n!}(x-a)^n+\frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1} \tag{1}$$

*holds true.*

*Proof.* We designate as $\varphi(x, a)$ the polynomial of degree $n$ with respect to $x$ which appears on the right-hand side of formula (1), i.e. set

$$\varphi(x, a)=f(a)+\frac{f'(a)}{1!}(x-a)+\frac{f''(a)}{2!}(x-a)^2+$$
$$\ldots+\frac{f^{(n)}(a)}{n!}(x-a)^n.$$

(It is known as Taylor's polynomial of degree $n$ for the function $f(x)$.)

Then we denote by $R_{n+1}(x)$ the difference

$$R_{n+1}(x)=f(x)-\varphi(x, a).$$

We shall prove the theorem if we establish that

$$R_{n+1}(x)=\frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}, \quad a<\xi<x.$$

We fix any value of $x$ from the indicated neighbourhood. For definiteness, we assume that $x>a$. We designate as $t$ the variable quantity, which varies on the interval $a\leqslant t\leqslant x$, and consider an auxiliary function

$$F(t)=f(x)-\varphi(x, t)-\frac{(x-t)^{n+1}R_{n+1}(x)}{(x-a)^{n+1}} \tag{2}$$

on the interval $[a, x]$.

On the interval $[a, x]$ the function $F(t)$ satisfies all the conditions of Rolle's theorem:

(1) it follows from relation (2) and from the conditions imposed on the function $f(x)$ that $F(t)$ is continuous and differentiable on $[a, x]$ since $f(t)$ and its derivatives up to the order $n$ inclusive are continuous and differentiable on $[a, x]$;

(2) setting $t=a$ in (2), we have

$$F(a)=f(x)-\varphi(x, a)-R_{n+1}(x)=R_{n+1}(x)-R_{n+1}(x)=0.$$

Setting $t=x$ in (2), we obtain

$$F(x)=f(x)-f(x)-\frac{f'(x)}{1!}(x-x)-\frac{f''(x)}{2!}(x-x)^2-$$
$$\ldots-\frac{f^{(n)}(x)}{n!}(x-x)^n-\frac{(x-x)^{n+1}R_{n+1}(x)}{(x-a)^{n+1}}=0.$$

Thus the condition $F(a) = F(x)$ is satisfied.

By virtue of Rolle's theorem, there is a point $\xi$ within the interval $[a, x]$ such that

$$F'(\xi) = 0. \tag{3}$$

We shall calculate the derivative $F'(t)$. Differentiating relation (2) with respect to $t$, we have

$$F'(t) = -f'(t) + \frac{f'(t)}{1!} - \frac{f''(t)}{1!}(x-t)$$
$$+ \frac{f''(t)}{2!} 2(x-t) - \frac{f'''(t)}{2!}(x-t)^2 +$$
$$\ldots + \frac{f^{(n)}(t)}{n!} n(x-t)^{n-1} - \frac{f^{(n+1)}(t)}{n!}(x-t)^n + \frac{(n+1)(x-t)^n R_{n+1}(x)}{(x-a)^{n+1}}.$$

It is easy to see that all the terms on the right-hand side of the relation, except for the last two, mutually cancel out. Thus

$$F'(t) = -\frac{f^{(n+1)}(t)}{n!}(x-t)^n + \frac{(n+1)(x-t)^n R_{n+1}(x)}{(x-a)^{n+1}}. \tag{4}$$

Setting $t = \xi$ in (4) and using relation (3), we obtain

$$F'(\xi) = -\frac{f^{(n+1)}(\xi)}{n!}(x-\xi)^n + \frac{(n+1)(x-\xi)^n R_{n+1}(x)}{(x-a)^{n+1}} = 0,$$

whence we have

$$R_{n+1}(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}. \blacksquare$$

Formula (1) is known as *Taylor's formula* and the expression for $R_{n+1}(x)$, as the *remainder in Lagrange's form*. We can rewrite it in another form. Since the point $\xi \in (a, x)$, there is a number $\theta$ from the interval $0 < \theta < 1$ such that $\xi = a + \theta(x - a)$, and the remainder assumes the form

$$R_{n+1}(x) = \frac{f^{(n+1)}[a+\theta(x-a)]}{(n+1)!}(x-a)^{n+1}, \quad 0 < \theta < 1.$$

This form of the remainder is most often used in applications.

**2. Another notation of Taylor's formula and of the remainder.** Taylor's formula (1) is often written in another form. We set $a = x_0$, $x - a = \Delta x$, $x = x_0 + \Delta x$ in (1) and then

$$f(x_0 + \Delta x) - f(x_0) = \frac{f'(x_0)}{1!}\Delta x + \frac{f''(x_0)}{2!}(\Delta x)^2 +$$
$$\ldots + \frac{f^{(n)}(x_0)}{n!}(\Delta x)^n + \frac{f^{(n+1)}(x_0 + \theta\Delta x)}{(n+1)!}(\Delta x)^{n+1}, \quad 0 < \theta < 1. \tag{5}$$

For $n = 0$ formula (5) yields Lagrange's formula

$$f(x_0 + \Delta x) - f(x_0) = f'(x_0 + \theta\Delta x)\Delta x.$$

We shall show that if the function $f^{(n+1)}(x)$ is bounded in the neighbourhood of the point $a$, then the remainder $R_{n+1}(x)$ is an infinitesimal of the order higher than that of $(x-a)^n$ as $x \to a$:

$$\lim_{x\to a} \frac{R_{n+1}(x)}{(x-a)^n} = \lim_{x\to a} \frac{f^{(n+1)}(\xi)}{(n+1)!} \frac{(x-a)^{n+1}}{(x-a)^n} = \lim_{x\to a} \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a) = 0$$

since the function $f^{(n+1)}(\xi)$ is bounded and $(x-a) \to 0$ as $x \to a$. Thus

$$R_{n+1}(x) = o\,[(x-a)^n] \text{ as } x \to a. \tag{6}$$

Formula (6) is known as the *remainder in Peano's* form.*

**3. Maclaurin's formula.** Maclaurin's** formula is Taylor's formula (1) for $a = 0$:

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \ldots + \frac{f^{(n)}(0)}{n!}x^n + R_{n+1}(x).$$

The remainder is

(1) in Lagrange's form $R_{n+1}(x) = \frac{f^{(n+1)}(\theta x)}{(n+1)!}x^{n+1}$, $0 < \theta < 1$;

(2) in Peano's form $R_{n+1}(x) = o(x^n)$.

**4. Expanding some elementary functions into Maclaurin's series.**

(1) $f(x) = e^x$. Since

$$f(x) = f'(x) = f''(x) = \ldots = f^{(n+1)}(x) = e^x,$$
$$f(0) = f'(0) = f''(0) = \ldots = f^{(n+1)}(0) = 1,$$

Maclaurin's formula has the form

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots + \frac{x^n}{n!} + o(x^n). \tag{7}$$

(2) $f(x) = \sin x$. Since

$$f^{(n)}(x) = \sin\left(x + n\frac{\pi}{2}\right), \quad f^{(n)}(0) = \sin\left(n\frac{\pi}{2}\right)$$
$$= \begin{cases} 0 & \text{for even } n, \\ (-1)^{(n-1)/2} & \text{for odd } n, \end{cases}$$

Maclaurin's formula has the form

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots + (-1)^{n-1}\frac{x^{2n-1}}{(2n-1)!} + o(x^{2n}). \tag{8}$$

---

* Peano, Giuseppe (1858-1932), an Italian mathematician and logician.
** Maclaurin, Colin (1698-1746), a Scottish mathematician.

(3) $f(x) = \cos x$. Since

$$f^{(n)}(x) = \cos\left(x + n\frac{\pi}{2}\right), \qquad f^{(n)}(0) = \cos\left(n\frac{\pi}{2}\right)$$
$$= \begin{cases} 0 & \text{for odd } n, \\ (-1)^{n/2} & \text{for even } n, \end{cases}$$

Maclaurin's formula has the form

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \ldots + (-1)^n \frac{x^{2n}}{(2n)!} + o(x^{2n+1}). \quad (9)$$

In formula (8) the remainder is written in the form $o(x^{2n})$ and not in the form $o(x^{2n-1})$ since the term following the last term is zero (the same refers to formula (9)).

(4) $f(x) = (1 + x)^\alpha$, where $\alpha$ is a real number. Since

$$f^{(n)}(x) = \alpha(\alpha - 1)\ldots(\alpha - n + 1)(1 + x)^{\alpha - n},$$
$$f^{(n)}(0) = \alpha(\alpha - 1)\ldots(\alpha - n + 1),$$

Maclaurin's formula has the form

$$(1+x)^\alpha = 1 + \frac{\alpha}{1!}x + \frac{\alpha(\alpha-1)}{2!}x^2 +$$
$$\ldots + \frac{\alpha(\alpha-1)\ldots(\alpha-n+1)}{n!}x^n + R_{n+1}(x),$$

where the remainder, in Lagrange's form, is equal to

$$R_{n+1}(x) = \frac{\alpha(\alpha-1)\ldots(\alpha-n)}{(n+1)!}(1+\theta x)^{\alpha-(n+1)}x^{n+1}, \quad 0 < \theta < 1.$$

In a special case when $\alpha = n$ is a natural number, $f^{(n+1)}(x) = 0$, and, consequently, $R_{n+1}(x) = 0$, we get a *binomial formula* known from elementary mathematics

$$(1+x)^n = 1 + \frac{n}{1!}x + \frac{n(n-1)}{2!}x^2 + \ldots + x^n. \quad (10)$$

The expansions given above show that using Maclaurin's formula, we can replace functions, with a definite degree of accuracy, by polynomials which are the basic elementary functions. It is easy to carry out arithmetic operations with polynomials or calculate the value of a polynomial at any point, etc. Taylor's and Maclaurin's formulas make it possible to approximate more complicated functions by polynomials. In addition, these formulas have a variety of applications. We shall consider here two of their applications.

**5. Using Maclaurin's formula to calculate limits.** Taylor's formula is an efficient means for calculating the limits of functions which we often encounter when we investigate the behaviour of functions.

**Example 1.** Find $\lim\limits_{x \to 0} \dfrac{\sin x - x}{x^3}$.

*Solution.* For $n=2$, we have from formula (8)

$$\lim_{x\to 0}\frac{x-\frac{x^3}{3!}+o(x^4)-x}{x^3}=\lim_{x\to 0}\frac{-\frac{1}{3!}+\frac{o(x^4)}{x^3}}{1}$$
$$=-\frac{1}{3!}+\lim_{x\to 0}\frac{o(x^4)}{x^3}=-\frac{1}{3!}+0=-\frac{1}{6}.$$

**Example 2.** Find $\lim\limits_{x\to 0}\dfrac{e^{-x^2/2}-\cos x}{x^3\sin x}$.

*Solution.* From formulas (7), (8) and (9) we have

$$\lim_{x\to 0}\frac{e^{-x^2/2}-\cos x}{x^3\sin x}=\lim_{x\to 0}\frac{1-x^2/2+x^4/8+o(x^4)-1+x^2/2-x^4/24}{x^3(x+o(x))}$$
$$=\lim_{x\to 0}\frac{x^4/8-x^4/24+o(x^4)}{x^4+o(x^4)}=\lim_{x\to 0}\frac{1/8-1/24+\alpha(x)}{1+\alpha(x)}=\frac{1}{8}-\frac{1}{24}=\frac{1}{12}$$

$\Big($the symbol $\alpha(x)$ denotes here the quantity $\frac{o(x^4)}{x^4}$ which is an infinitesimal as $x\to 0\Big)$.

**6. Calculating the number e.** In Item 2 of 1.2.3 we introduced the number e as the limit of the sequence $\{(1+1/n)^n\}$ and got a rough estimate $2\leqslant e\leqslant 3$.

We shall show how the number e can be calculated with any degree of accuracy. For that purpose, we write formula (7) with the remainder in Lagrange's form:

$$e^x=1+\frac{x}{1!}+\frac{x^2}{2!}+\ldots+\frac{x^n}{n!}+\frac{e^{\theta x}}{(n+1)!}x^{n+1},\quad 0<\theta<1. \tag{11}$$

If we replace the function $e^x$ by its Taylor's polynomial of degree $n$, we get an approximate equality

$$e^x\cong 1+\frac{x}{1!}+\frac{x^2}{2!}+\ldots+\frac{x^n}{n!} \tag{12}$$

whose absolute error is

$$|R_{n+1}(x)|=\frac{e^{\theta x}}{(n+1)!}|x|^{n+1},\quad 0<\theta<1.$$

If we consider the function $e^x$ for $-1\leqslant x\leqslant 1$, then

$$|R_{n+1}(x)|\leqslant\frac{e^{\theta x}}{(n+1)!}<\frac{3}{(n+1)!}. \tag{13}$$

Setting $x=1$ in (12), we get an approximate value of the number e:

$$e\cong 1+1+\frac{1}{2!}+\ldots+\frac{1}{n!}.$$

Here the absolute error is smaller than $3/(n+1)!$ If we have to calculate the value of e with an accuracy to within 0.001, then we must use the inequality $3/(n+1)! < 0.001$, or $(n+1)! > 3000$ to calculate the number $n$. Consequently, if we take $n = 6$, then the required inequality is satisfied.

Thus, using Maclaurin's formula, we can calculate the number e with any degree of accuracy, and in that case, the algorithm for calculating the number e, based on formulas (11) and (13), is easily realized on computers.

### 1.6.4. Investigating the Behaviour of Functions and Constructing Graphs

**1. Criterion of monotonicity of a function.**

**Theorem 6.7.** *If the function $f(x)$ is differentiable on the interval $(a, b)$ and $f'(x) \geqslant 0$ ($f'(x) \leqslant 0$) on $(a, b)$, then the function $f(x)$ does not decrease (does not increase) on $(a, b)$.*

*Proof.* For the sake of definiteness, we shall consider the case $f'(x) \geqslant 0$. Let $x_1$ and $x_2$ be two arbitrary points from $(a, b)$ and

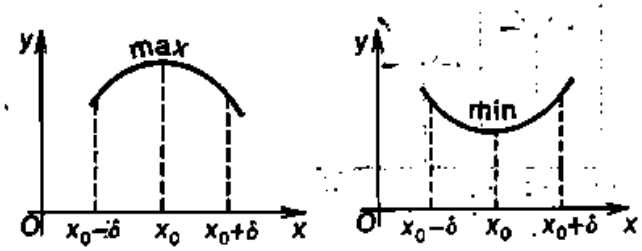



Fig. 76

$x_1 < x_2$. Then all the conditions of Lagrange's theorem are fulfilled on the interval $[x_1, x_2]$, according to which we have

$$f(x_2) - f(x_1) = f'(c)(x_2 - x_1), \quad c \in (x_1, x_2).$$

By the hypothesis, $f'(c) \geqslant 0$, $x_2 - x_1 > 0$, and therefore $f(x_2) - f(x_1) \geqslant 0$, or $f(x_2) \geqslant f(x_1)$, i.e. the function $f(x)$ does not decrease on $(a, b)$.

The proof for the case $f'(x) \leqslant 0$ is analogous. ■

**Remark.** We can prove in the same way that if $f'(x) > 0$ ($f'(x) < 0$) on $(a, b)$, then $f(x)$ increases (decreases) on $(a, b)$.

**2. Finding the points of local extremum of a function.**

**Definition.** *The point $x_0$ is a point of a strict local maximum (minimum) of the function $f(x)$ if for all $x$ belonging to a certain $\delta$-neighbourhood of the point $x_0$ the inequality $f(x) < f(x_0)$ ($f(x) > f(x_0)$) is satisfied for $x \neq x_0$* (Fig. 76).

The name that is used both for a local maximum (max) and a local minimum (min) is a *local extremum.*

It follows from the definition that the notion of extremum is local in nature in the sense that the inequality $f(x) < f(x_0)$ ($f(x) > f(x_0)$) may not be satisfied for all values of $x$ in the domain of definition of a function but must be satisfied in some neighbourhood of the point $x_0$. A function can evidently have several local maxima and several local minima, and it may so happen that a local maximum turns out to be smaller than a local minimum.

**Theorem 6.8 (the necessary condition for a local extremum).** *If the function $f(x)$ has a local extremum at the point $x_0$ and is differentiable at this point, then $f'(x_0) = 0$.*

*Proof.* Since the function $f(x)$ has a local extremum at the point $x_0$, there is an interval $(x_0 - \delta, x_0 + \delta)$ in which the value $f(x_0)$ is the

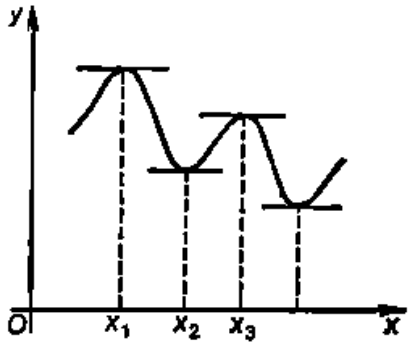



Fig. 77

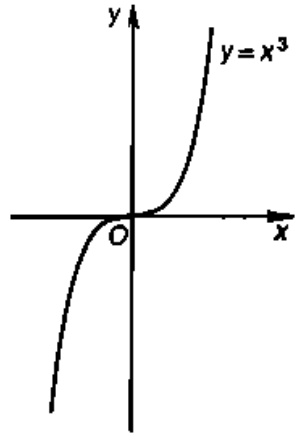



Fig. 78

greatest or the least of all other values of this function. Then, by Theorem 6.1, the derivative of the function at the point $x_0$ is zero, i.e. $f'(x_0) = 0$. ■

Theorem 6.8 has the following geometric meaning. If $x_1$, $x_2$, $x_3$ are points of local extremum and there are tangents at the corresponding points of the graph, then these tangents are parallel to the $x$-axis (Fig. 77).

Points of this kind are sometimes called stationary points. We shall call them *points of a possible extremum*. If $x_0$ is a point of a possible extremum, i.e. $f'(x_0) = 0$, then it may not be a point of a local maximum or minimum. For example, if $f(x) = x^3$, then $f'(x) = 3x^2 = 0$ for $x = 0$, but nevertheless there is no local extremum at the point $x = 0$ (Fig. 78). We shall establish the sufficient condition for the existence of a local extremum. The following theorem is devoted to this.

**Theorem 6.9 (the sufficient condition for a local extremum).** *Let the function $f(x)$ be differentiable in a certain $\delta$-neighbourhood of the point $x_0$. Then, if $f'(x) > 0$ ($f'(x) < 0$) for all $x$ from $(x_0 - \delta, x_0)$ and $f'(x) < 0$ ($f'(x) > 0$) for all $x$ from $(x_0, x_0 + \delta)$, then the func-*

*tion $f(x)$ has a local maximum (minimum) at the point $x_0$, but if $f'(x)$ has one and the same sign in the whole $\delta$-neighbourhood of the point $x_0$, then there is no local extremum at the point $x_0$.*

In other words, if $f'(x)$ changes sign from "+" to "—", when passing through the point $x_0$, then $x_0$ is a point of a local maximum; if $f'(x)$ changes sign from "—" to "+" at the point $x_0$, then $x_0$ is a point of a local minimum; now if $f'(x)$ does not change sign at the point $x_0$, then there is no extremum at the point $x_0$.

*Proof.* Assume that $f'(x)$ changes sign from "+" to "—" when passing through the point $x_0$ and let $x \in (x_0 - \delta, x_0)$. We apply Lagrange's formula to the function $f(x)$ on the interval $[x, x_0]$ and get

$$f(x_0) - f(x) = f'(c)(x - x), \ c \in (x, x_0).$$

Since $f'(x) > 0$ on $(x_0 - \delta, x_0)$, it follows that $f'(c) > 0$, and, in addition, $x_0 - x > 0$ and, consequently,

$$f(x_0) - f(x) > 0, \text{ or } f(x_0) > f(x). \tag{1}$$

Let us consider now the case when $x \in (x_0, x_0 + \delta)$. We apply Lagrange's formula to the function $f(x)$ on the interval $[x_0, x]$ and obtain

$$f(x) - f(x_0) = f'(c)(x - x_0), \ c \in (x_0, x).$$

Since $f'(x) < 0$ on $(x_0, x_0 + \delta)$, it follows that $f'(c) < 0$, and, in addition, $x - x_0 > 0$ and, consequently,

$$f(x) - f(x_0) < 0, \text{ or } f(x_0) > f(x). \tag{2}$$

It follows from (1) and (2) that the inequality $f(x) < f(x_0)$ is satisfied for $x \neq x_0$ in the neighbourhood of the point $x_0$ being considered, and this means that the function $f(x)$ has a local maximum at the point $x_0$.

The case when the function $f'(x)$ changes sign from "—" to "+" can be treated by analogy.

It remains to consider the case when $f'(x)$ does not change sign. Let $f'(x) > 0$ in some neighbourhood of the interval $(x_0 - \delta, x_0 + \delta)$. Then, according to Theorem 6.7, the function $f(x)$ does not decrease on $(x_0 - \delta, x_0 + \delta)$, i.e. the inequality $f(x) < f(x_0)$ is satisfied for any $x < x_0$ and the inequality $f(x) > f(x_0)$ is satisfied for any $x > x_0$. This means that the point $x_0$ is not a point of a local extremum. ■

**Remark.** Theorem 6.9 remains true when the function $f(x)$ is not differentiable at the point $x_0$ itself but is only continuous. Thus, for instance, the function $f(x) = |x|$ is continuous at the point $x = 0$ but is not differentiable.

We shall discuss, for example, the problem of finding points of a local extremum of the function $f(x) = x^3 - 3x$. We find the derivative: $f'(x) = 3x^2 - 3 = 3(x^2 - 1)$. Solving the equation

$3(x^2-1)=0$, we get two points of a possible extremum: $x_1=-1$ and $x_2=1$. It is convenient to continue the investigation using an auxiliary drawing (Fig. 79). Marking the points $x_1=-1$ and $x_2=1$ on it and investigating the sign of $f'(x)$ in the neighbourhood

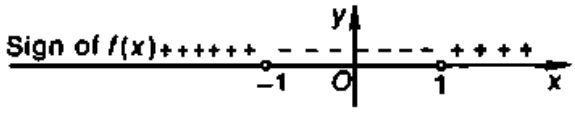



Fig. 79

of these points, we find that $f(x)$ has a local maximum at the point $x_1=-1$ and a local minimum at the point $x_2=1$. Then we find that $y_{max}=f(-1)=2$ and $y_{min}=f(1)=-2$.

Figure 79 also shows the intervals of monotonicity of $f(x)$: $(-\infty, -1)$, $(-1, 1)$, $(1, +\infty)$, in the first and third of which the function increases and in the second decreases.

**3. The direction of convexity and the inflection points of the graph of a function.** Let the function $y=f(x)$ be differentiable on the

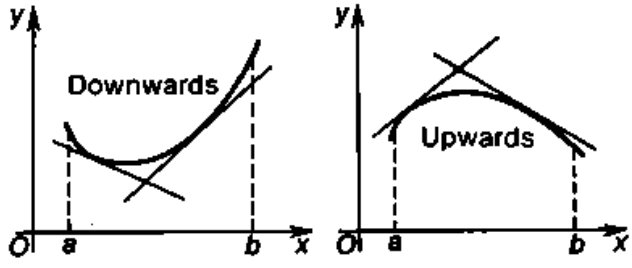



Fig. 80

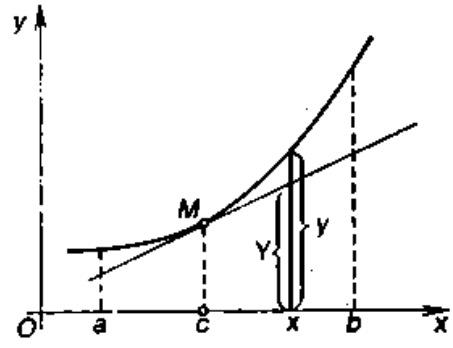



Fig. 81

interval $(a, b)$. Then there is a tangent to the graph of the function $y=f(x)$ at any point $M(x, f(x))$ of the graph $(a<x<b)$, and the tangent is not parallel to the $y$-axis since its slope, equal to $f'(x)$, is finite.

**Definition 1.** *We say that the graph of the function $y=f(x)$ is convex downwards (upwards) on $(a, b)$ if it lies not lower (not higher) than any tangent drawn to the graph of the function on $(a, b)$* (Fig. 80).

**Theorem 6.10.** *If the function $y=f(x)$ has a second derivative on the interval $(a, b)$ and $f''(x)\geqslant 0$ ($f''(x)\leqslant 0$) at all points of $(a, b)$, then the graph of the function $y=f(x)$ is convex downwards (upwards) on $(a, b)$.*

*Proof.* For the sake of definiteness, we shall consider the case when $f''(x)\geqslant 0$ on $(a, b)$. We designate an arbitrary point of $(a, b)$ as $c$ (Fig. 81). We must prove that the graph of the function $y=f(x)$

lies not lower than the tangent which passes through the point $M(c, f(c))$.

We write the equation of this tangent designating the running ordinate of its points as $Y$: $Y - f(c) = f'(c)(x - c)$, or

$$Y = f(c) + f'(c)(x - c). \tag{3}$$

We expand the function $f(x)$ in the neighbourhood of the point $c$ using Taylor's formula for $n = 1$. We get

$$y = f(x) = f(c) + \frac{f'(c)}{1!}(x - c) + \frac{f''(\xi)}{2!}(x - c)^2, \quad \xi \in (c, x). \tag{4}$$

Relation (4) is valid for any $x$ from $(a, b)$.

Subtracting relation (3) from (4), we have

$$y - Y = \frac{f''(\xi)}{2!}(x - c)^2. \tag{5}$$

Since, by the hypothesis, $f''(x) \geqslant 0$ on $(a, b)$, the right-hand side of (5) is nonnegative, i.e. $y - Y \geqslant 0$ for all $x$ from $(a, b)$ or $y \geqslant Y$.

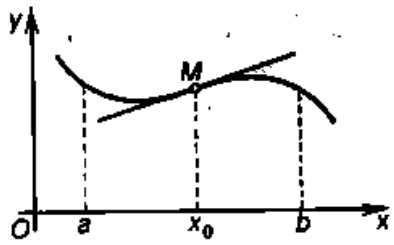



Fig. 82

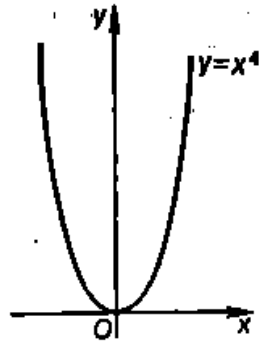



Fig. 83

The last inequality proves that everywhere within $(a, b)$ the graph of the function $y = f(x)$ lies not lower than tangent (3).

For the case when $f''(x) \leqslant 0$ the proof of the theorem is similar. ■

**Definition 2.** *The point $M(x_0, f(x_0))$ is an inflection point of the graph of the function $y = f(x)$ if at the point $M$ the graph has a tangent and there is a neighbourhood of the point $x_0$ in whose interior the graph of the function $y = f(x)$ has convexities of different directions on the left and on the right of the point $x_0$.*

Evidently, at a point of inflection the tangent cuts the graph of the function since on one side of that point the graph lies below the tangent and on the other side it lies above it, i.e. in the neighbourhood of the inflection point the graph of the function geometrically passes from one side of the tangent to the other and "bends" over it (Fig. 82).

**Theorem 6.11 (the necessary condition for an inflection point).** *Assume that the graph of the function $y = f(x)$ has an inflection at the*

*point $M(x_0, f(x_0))$ and let the function $y = f(x)$ possess a continuous second derivative at the point $x_0$. Then $f''(x)$ vanishes at the point $x_0$, i.e. $f''(x_0) = 0$.*

*Proof.* We assume the contrary, i.e. let $f''(x_0) \neq 0$. Then, by virtue of the continuity of the second derivative, according to Theorem 4.8 on the constancy of sign of a continuous function, there is a neighbourhood of the point $x_0$ in which $f''(x) < 0$ ($f''(x) > 0$) and, hence, according to Theorem 6.10, the graph of the function $y = f(x)$ has a definite direction of convexity in this neighbourhood. But this contradicts the existence of inflection at the point $M(x_0, f(x_0))$ (see Fig. 82). The contradiction obtained proves the theorem. ■

It should be pointed out that not every point $M(x_0, f(x_0))$, for which $f''(x_0) = 0$, is an inflection point. For example, the graph of the function $f(x) = x^4$ has no inflection at the point (0, 0), although $f''(x) = 12x^2 = 0$ for $x = 0$ (Fig. 83). Therefore the equality of the second derivative to zero is only the necessary condition for inflection. The points $M(x_0, f(x_0))$ of the graph, for which $f''(x_0) = 0$, are called *critical*, or *singular*, *points*. We must also investigate the problem of the existence of inflection at every critical point, for which purpose we must establish the sufficient condition for inflection.

**Theorem 6.12 (the sufficient condition for an inflection point).** *Let the function $y = f(x)$ possess a second derivative in some neighbourhood of the point $x_0$. If in the interior of the indicated neighbourhood $f''(x)$ has different signs on the left and on the right of the point $x_0$, then the graph of the function $y = f(x)$ has an inflection at the point $M(x_0, f(x_0))$.*

*Proof.* From the fact that $f''(x)$ has different signs on the left and on the right of the point $x_0$ we infer, by virtue of Theorem 6.10, that the direction of convexity of the graph of the function on the left and on the right of the point $x_0$ is different. And this signifies the existence of inflection at the point $M(x_0, f(x_0))$. ■

**Remark.** The theorem remains true if $f(x)$ has a second derivative in some neighbourhood of the point $x_0$, except for the point $x_0$ itself, and there is a tangent to the graph of the function at the point $M$. If in the interior of the indicated neighbourhood $f''(x)$ has different signs on the left and on the right of the point $x_0$, then the graph of the function $y = f(x)$ has an inflection at the point $M(x_0, f(x_0))$. The proof of this fact is similar to the proof of the theorem.

Consider an example: $f(x) = x^{1/3}$. This function has an infinite derivative at the point $x = 0$, and the tangent to the graph of the function at the point $O(0, 0)$ coincides with the $y$-axis. The second derivative at the point $x = 0$ does not exist. However, the graph of the function $y = x^{1/3}$ has an inflection at the point $O(0, 0)$ since the second derivative $f''(x) = 2/(9x^{5/3})$ has different signs on the left and on the right of the point $x = 0$ (Fig. 84).

Thus the second derivative can be used to investigate the direction of convexity and inflection points of the graph of a function. For example, we can consider the function $f(x) = x^3 - 3x$ which we discussed in Item 2. We shall mark the sign of the second derivative on the auxiliary drawing shown in Fig. 79. We find the second derivative: $f''(x) = 6x$. From the equation $6x = 0$ we find one critical point, $O(0, 0)$. Having marked the point $x = 0$ on the auxiliary drawing (Fig. 85) and investigated the sign of $f''(x)$ in its neighbourhood, we find that to the left of the point $x = 0$ the derivative $f''(x) < 0$ (the graph is convex upwards) and to the right of the point $x = 0$ the derivative $f''(x) > 0$ (the graph is convex downwards), i.e. the point $O(0, 0)$ is an inflection point of the graph of the function in question. This graph is shown diagrammatically in Fig. 86.

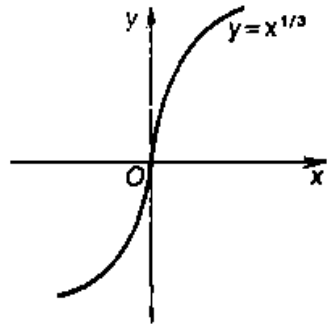


Fig. 84

Let us now prove that the part of the ellipse $\left(\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1\right)$, which lies in the upper half-plane ($y \geqslant 0$), is convex upwards on the inter-

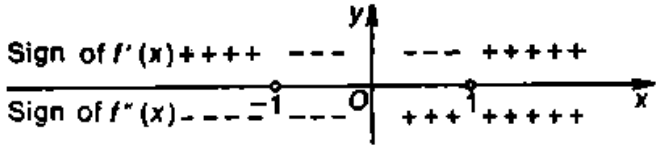


Fig. 85

val $(-a, a)$. Indeed, from the equation of the ellipse we have $y = \frac{b}{a}\sqrt{a^2 - x^2}$. Next we find that $y' = -\frac{b}{a}\frac{x}{\sqrt{a^2 - x^2}}$, $y'' = -\frac{ba}{(a^2 - x^2)^{3/2}}$. It follows from the expression for the second derivative that it is negative on the interval $(-a, a)$. This means that the given curve is convex upwards throughout the interval $(-a, a)$ (see Fig. 33).

We can show by analogy (do it yourself!) that the part of the hyperbola $\left(\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1\right)$, which lies in the upper half-plane, is convex upwards on the intervals, $(a, +\infty)$ and $(-\infty, -a)$.

**4. The asymptotes to the graph of a function.** When we investigate the behaviour of a function at infinity, i.e. as $x \to +\infty$ and $x \to -\infty$, or in the vicinity of the points of discontinuity of the second kind, we often find that the graph of the function comes arbitrarily

close to a certain straight line. The straight lines of this kind are known as *asymptotes**.

There are three kinds of an asymptote: *vertical*, *horizontal* and *oblique*.

**Definition 1.** *The straight line $x = x_0$ is a vertical asymptote to the graph of the function $y = f(x)$ if at least one of the limiting values $\lim\limits_{x \to x_0+} f(x)$ or $\lim\limits_{x \to x_0-} f(x)$ is equal to $+\infty$ or $-\infty$.*

In this case the distance from the point $M(x, f(x))$ to the straight line $x = x_0$ is $d = \sqrt{(x - x_0)^2 + (f(x) - f(x))^2} = |x - x_0|$ and,

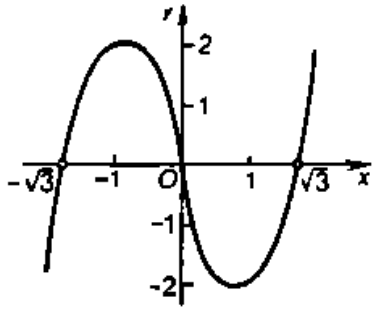



Fig. 86

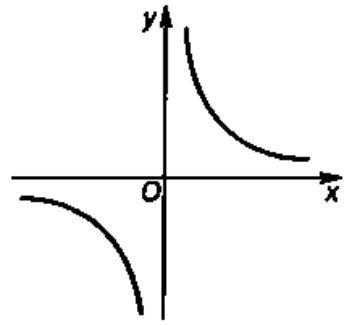



Fig. 87

consequently, $d \to 0$ as $x \to x_0$. For instance, the graph of the function $y = f(x) = 1/x$ (Fig. 87) has a vertical asymptote $x = 0$ since $f(x) \to +\infty$ as $x \to 0+$ and $f(x) \to -\infty$ as $x \to 0-$.

**Definition 2.** *The straight line $y = A$ is a horizontal asymptote to the graph of the function $y = f(x)$ as $x \to +\infty$ $(x \to -\infty)$ if $\lim\limits_{\substack{x \to +\infty \\ (x \to -\infty)}} f(x) = A$.*

In this case the distance from the point $M(x, f(x))$ to the straight line $y = A$ is $d = \sqrt{(x - x)^2 + (f(x) - A)^2} = |f(x) - A|$ and, consequently, $d \to 0$ as $x \to \infty$ since $\lim\limits_{x \to \infty} |f(x) - A| = 0$. For instance, the graph of the function $y = 1/x$ considered above has a horizontal asymptote $y = 0$ as $x \to +\infty$ and as $x \to -\infty$ since $1/x \to 0$ as $x \to +\infty$ and as $x \to -\infty$.

**Definition 3.** *The straight line $y = kx + b$ $(k \neq 0)$ is an oblique asymptote to the graph of the function $y = f(x)$ as $x = +\infty$ $(x \to -\infty)$ if the function $f(x)$ can be represented as*

$$f(x) = kx + b + \alpha(x), \tag{6}$$

*where $\alpha(x) \to 0$ as $x \to +\infty$ $(x \to -\infty)$.*

Let us discuss the geometric meaning of the oblique asymptote. For the sake of definiteness, we shall consider the case when $x \to +\infty$. (The case when $x \to -\infty$ can be treated by analogy.)

* We have come across the notion of an asymptote in analytic geometry when we considered the hyperbola (see 1.3.6, Item 2).

Assume that $M(x, y)$ is a point of the graph of the function $y = f(x)$ and let the straight line $\tilde{y} = kx + b$ be an oblique asymptote to the graph of the function as $x \to +\infty$. We designate the running ordinate of the point on the asymptote as $\tilde{y}$ and the point on the asymptote as $N(x, \tilde{y})$ (Fig. 88). Then $|MN| = |y - \tilde{y}| = |f(x) - (kx + b)| = |\alpha(x)| \to 0$ as $x \to +\infty$. We drop a perpendicular $MP$ from the point $M$ to the asymptote. The distance $d$ from the point $M$ to the asymptote is $|MP| = |MN| \cos \alpha$, where $\alpha$ is the angle between the asymptote and the $x$-axis and, consequently, $\lim\limits_{x \to +\infty} d = 0$.

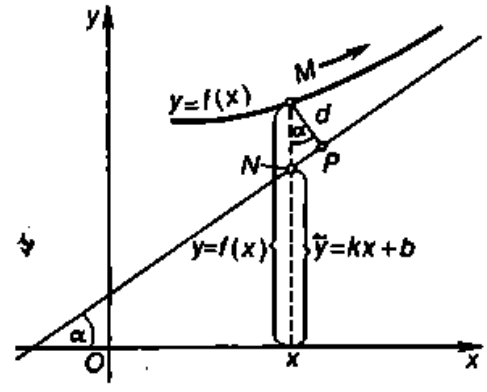


Fig. 88

Thus the distance from the point $M(x, y)$ of the graph of the function to the asymptote tends to zero as $x \to +\infty$, i.e. the graph of the function approaches the asymptote indefinitely as $x \to +\infty$.

We shall consider the way of finding an oblique asymptote, i.e. the way of determining the numbers $k$ and $b$ in the equation of the asymptote. Dividing relation (6) by $x$ and passing to the limit as $x \to +\infty$, we obtain

$$\lim_{x \to +\infty} \frac{f(x)}{x} = \lim_{x \to +\infty} \left[ k + \frac{b}{x} + \frac{\alpha(x)}{x} \right] = k.$$

Thus

$$k = \lim_{x \to +\infty} \frac{f(x)}{x}. \tag{7}$$

Furthermore, from relation (6) we get $\lim\limits_{x \to +\infty} [f(x) - kx] = \lim\limits_{x \to +\infty} [b + \alpha(x)] = b$. Thus

$$b = \lim_{x \to +\infty} [f(x) - kx]. \tag{8}$$

We have proved that if the straight line $\tilde{y} = kx + b$ is an oblique asymptote, then the numbers $k$ and $b$ can be found from relations (7) and (8). Conversely, if both limits (7) and (8) exist and $k \neq 0$, then the straight line $\tilde{y} = kx + b$ is an oblique asymptote to the graph of the function $y = f(x)$ as $x \to +\infty$.

Indeed, setting $\alpha(x) = f(x) - kx - b$ and using relation (8), we find that $\lim\limits_{x \to +\infty} \alpha(x) = 0$. Consequently, relation (6) holds true: $f(x) = kx + b + \alpha(x)$, where $\lim\limits_{x \to +\infty} \alpha(x) = 0$, i.e. the

straight line $\tilde{y} = kx + b$ is an oblique asymptote to the graph of the function as $x \to +\infty$.

It is practically expedient to seek the asymptotes in the following order: (1) vertical asymptotes, (2) horizontal asymptotes, (3) oblique asymptotes.

**Example 1.** Find the asymptotes to the graph of the function $y = \frac{x^2+2x-3}{x}$.

*Solution.* (1) We seek vertical asymptotes. The point $x = 0$ is a point of discontinuity of the second kind of the given function, with $y \to +\infty$ as $x \to 0-$ and $y \to -\infty$ as $x \to 0+$. Consequently, the axis of ordinates ($x = 0$) is a vertical asymptote.

(2) We seek horizontal asymptotes:

$$\lim_{\substack{x\to+\infty\\(x\to-\infty)}} \frac{x^2+2x-3}{x} = \lim_{\substack{x\to+\infty\\(x\to-\infty)}} \left(x+2-\frac{3}{x}\right) = +\infty \quad (-\infty).$$

Consequently, there are no horizontal asymptotes.

(3) We seek oblique asymptotes:

$$k = \lim_{\substack{x\to+\infty\\(x\to-\infty)}} \frac{f(x)}{x} = \lim_{\substack{x\to+\infty\\(x\to-\infty)}} \left(1+\frac{2}{x}-\frac{3}{x^2}\right) = 1,$$

$$b = \lim_{\substack{x\to+\infty\\(x\to-\infty)}} [f(x)-kx] = \lim_{\substack{x\to+\infty\\(x\to-\infty)}} \left(\frac{x^2+2x-3}{x} - x\right)$$

$$= \lim_{\substack{x\to+\infty\\(x\to-\infty)}} \left(\frac{2x-3}{x}\right) = \lim_{\substack{x\to+\infty\\(x\to-\infty)}} \left(2-\frac{3}{x}\right) = 2.$$

Consequently, the straight line $y = x + 2$ is an oblique asymptote to the graph of the given function both as $x \to +\infty$ and as $x \to -\infty$.

The graph of the function is shown diagrammatically in Fig. 89.

**Example 2.** Prove that the hyperbola $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$ has the straight lines $y = \pm\frac{b}{a}x$ as its oblique asymptotes.

*Solution.* Since $y = \pm\frac{b}{a}\sqrt{x^2-a^2}$, it follows that

$$\lim_{\substack{x\to+\infty\\(x\to-\infty)}} \frac{f(x)}{x} = \lim_{\substack{x\to+\infty\\(x\to-\infty)}} \left[\pm\frac{b}{a}\sqrt{1-\left(\frac{a}{x}\right)^2}\right] = \pm\frac{b}{a},$$

$$\lim_{\substack{x\to+\infty\\(x\to-\infty)}} [f(x)-kx] = \lim_{\substack{x\to+\infty\\(x\to-\infty)}} \left(\pm\frac{b}{a}\sqrt{x^2-a^2} \mp \frac{b}{a}x\right)$$

$$= \lim_{\substack{x\to+\infty\\(x\to-\infty)}} \left[\pm\frac{b}{a}\left(\sqrt{x^2-a^2}-x\right)\right] = \pm\frac{b}{a} \lim_{\substack{x\to+\infty\\(x\to-\infty)}} \frac{-a^2}{\sqrt{x^2-a^2}+x^2} = 0.$$

Consequently, the straight lines $y = \pm \frac{b}{a} x$ are oblique asymptotes to the hyperbola both as $x \to +\infty$ and as $x \to -\infty$.

**5. The scheme of investigating the graph of a function.** Let us consider a model scheme which it is expedient to use when investigating the behaviour of a function and constructing its graph.

1. Find the domain of definition of the function.
2. Find the points of intersection of the graph of the function with the coordinate axes.
3. Find the asymptotes.
4. Find the points of a possible extremum.
5. Find the critical points.
6. Using an auxiliary drawing, investigate the sign of the first and the second derivative. Find the intervals of increase and decrease of the function, determine the direction of convexity of the graph, the points of extremum and of inflection.
7. Construct the graph taking into account the investigation carried out in Items 1-6.

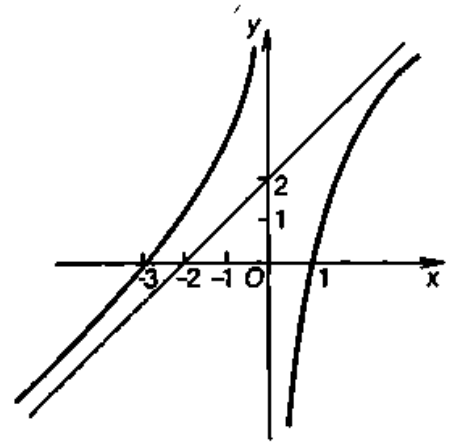


Fig. 89

It is expedient to verify at the beginning of the investigation whether the given function is even or odd in order to use the symmetry of the graph with respect to the axis of ordinates or to the origin when constructing the graph.

By way of example, we shall use the scheme given above to construct the graph of the function $y = (x^2 + 1)/(x - 1)$.

1. The domain of definition of the function is the set of all real numbers, except for $x = 1$ for which the denominator vanishes.

2. Since the equation $x^2 + 1 = 0$ does not have real roots, the graph of the function does not possess points of intersection with the $x$-axis but cuts the $y$-axis at the point $(0, -1)$.

3. We seek asymptotes. We investigate the behaviour of the function in the vicinity of the point of discontinuity $x = 1$. Since $y \to -\infty$ as $x \to 1-$ and $y \to +\infty$ as $x \to 1+$, the straight line $x = 1$ is a vertical asymptote to the graph of the function.

If $x \to +\infty$ $(x \to -\infty)$, then $y \to +\infty$ $(y \to -\infty)$ and, consequently, the graph has no horizontal asymptote. Furthermore, the existence of limits

$$k = \lim_{\substack{x \to +\infty \\ (x \to -\infty)}} \frac{f(x)}{x} = \lim_{\substack{x \to +\infty \\ (x \to -\infty)}} \frac{x^2+1}{x^2-x} = \lim_{\substack{x \to +\infty \\ (x \to -\infty)}} \frac{1+1/x^2}{1-1/x} = 1,$$

$$b = \lim_{\substack{x\to+\infty \\ (x\to-\infty)}} [f(x) - kx] = \lim_{\substack{x\to+\infty \\ (x\to-\infty)}} \left(\frac{x^2+1}{x-1} - x\right)$$

$$= \lim_{\substack{x\to+\infty \\ (x\to-\infty)}} \frac{1+x}{x-1} = \lim_{\substack{x\to+\infty \\ (x\to-\infty)}} \frac{1+1/x}{1-1/x} = 1$$

implies that both as $x \to +\infty$ and as $x \to -\infty$ the graph of the function has an oblique asymptote $y = x + 1$.

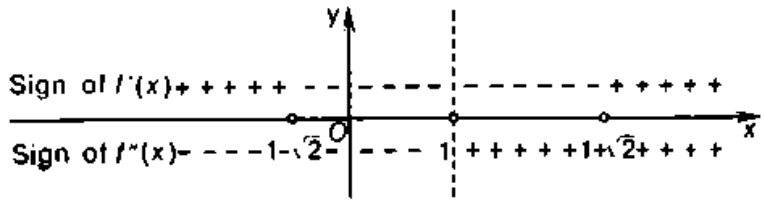



Fig. 90

4. To find the points of a possible extremum, we calculate the first derivative of the function:

$$f'(x) = \frac{2x(x-1)-(x^2+1)}{(x-1)^2} = \frac{2x^2-2x-x^2-1}{(x-1)^2} = \frac{x^2-2x-1}{(x-1)^2}.$$

Solving the equation $x^2 - 2x - 1 = 0$, we get two points of a possible extremum: $x_1 = 1 - \sqrt{2}$ and $x_2 = 1 + \sqrt{2}$.

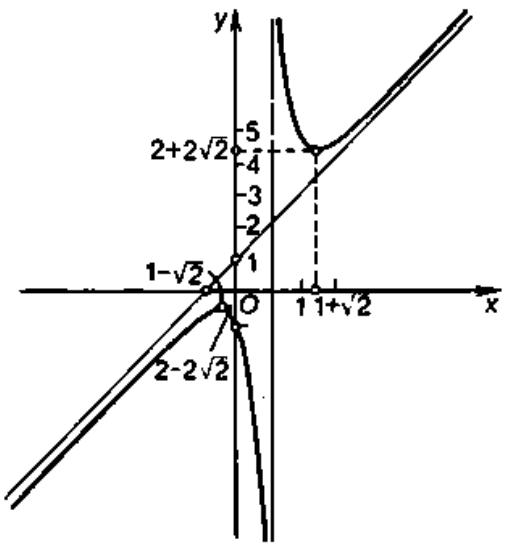



Fig. 91

5. To find the critical points, we calculate the second derivative:

$$f''(x) = \frac{(2x-2)(x-1)^2 - 2(x-1)(x^2-2x-1)}{(x-1)^4} = \frac{4}{(x-1)^3}.$$

Since $f''(x)$ does not vanish, there are no critical points.

6. We make an auxiliary drawing and investigate the sign of the first and the second derivative (Fig. 90). We find that the function increases on $(-\infty, 1 - \sqrt{2})$, decreases on $(1 - \sqrt{2}, 1 + \sqrt{2})$ and again increases on $(1 + \sqrt{2}, +\infty)$. The points of extremum are: (1) maximum for $x = 1 - \sqrt{2}$, with $f(1 - \sqrt{2}) = 2 - 2\sqrt{2}$; (2) minimum for $x = 1 + \sqrt{2}$, with $f(1 + \sqrt{2}) = 2 + 2\sqrt{2}$. On $(-\infty, 1)$ the graph is convex upwards and on $(1, +\infty)$ it is convex downwards.

7. Using the data obtained, we construct the scheme of the graph (Fig. 91).

### 1.6.5. Interpolation of Functions

Interpolation is used in solving many theoretical and applied problems which involve calculations.

**1. Formulation of the problem.** Let the values of the function $y = f(x)$ be defined on the interval $[a, b]$ at the points $a \leqslant x_0 < x_1 < x_2 < \ldots < x_n \leqslant b$:

$$f(x_0) = y_0,\ f(x_1) = y_1,\ f(x_2) = y_2,\ \ldots,\ f(x_n) = y_n.$$

We have to find a polynomial of a degree not higher than $n$:

$$P_n(x) = a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \ldots + a_{n-1}x + a_n, \tag{1}$$

which assumes, at the points $x_0, x_1, \ldots, x_n$, the same values as the given function, i.e. the relations

$$P_n(x_i) = f(x_i) = y_i,\ i = 0, 1, 2, \ldots, n, \tag{2}$$

are satisfied. In other words, we have to find a polynomial of form (1) which on the interval $[a, b]$ would be an approximation of the function $y = f(x)$. The problem formulated is known as an *interpolation problem*, polynomial (1) is an *interpolation polynomial*, and the points $x_0, x_1, \ldots, x_n$ are *interpolation nodes*. The solution of this problem makes it possible to find approximate values of the function $f(x)$ at the points $x$ which lie between the nodes. This is important when the function is defined only at the points $x_0, x_1, \ldots, x_n$ and we have to find its values in the intervals between those points and when the function $f(x)$ is defined by a formula on the whole interval $[a, b]$ but its calculation with the use of this formula is very labour-consuming.

We shall show that there always exists an interpolation polynomial of form (1) which satisfies conditions (2) and that polynomial is unique. For simplicity, we shall only consider the case when $n = 2$, i.e. the case of a second-degree polynomial

$$P_2(x) = a_0x^2 + a_1x + a_2. \tag{3}$$

Successively substituting the numbers $x_0, x_1, x_2$ for $x$ in equation (3) and bearing in mind that at these points the polynomial assumes the values $y_0, y_1, y_2$ respectively, we get a system of three equations of the first degree with three unknown coefficients $a_0, a_1, a_2$:

$$\begin{cases} a_0x_0^2 + a_1x_0 + a_2 = y_0, \\ a_0x_1^2 + a_1x_1 + a_2 = y_1, \\ a_0x_2^2 + a_1x_2 + a_2 = y_2. \end{cases}$$

Since the numbers $x_0$, $x_1$, $x_2$ are different, the determinant of this system is nonzero:

$$\begin{vmatrix} x_0^2 & x_0 & 1 \\ x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \end{vmatrix} = (x_1 - x_0)(x_2 - x_0)(x_1 - x_2) \neq 0.$$

Consequently (see 1.10.3), the solution of this system exists and is unique, and this proves the statement of the problem. In terms of geometry, this means that a single curve defined by equation (3) passes through three points $M_0(x_0, y_0)$, $M_1(x_1, y_1)$, $M_2(x_2, y_2)$. Thus the interpolation polynomial (1) always exists and is unique. Later on we shall discuss various forms of an interpolation polynomial.

**2. Lagrange's formula for interpolation.** We shall consider the problem of seeking the coefficients of the interpolation polynomial (1). Substituting this polynomial into system (2), we get a system of $n+1$ first-degree equations with $n+1$ unknown coefficients $a_0, a_1, \ldots, a_n$:

$$\begin{cases} a_0 + a_1 x_0 + \ldots + a_n x_0^n = y_0, \\ a_0 + a_1 x_1 + \ldots + a_n x_1^n = y_1, \\ \cdots\cdots\cdots\cdots\cdots \\ a_0 + a_1 x_n + \ldots + a_n x_n^n = y_n, \end{cases}$$

solving which, we find the values of the coefficients $a_0, a_1, \ldots, a_n$. Substituting these values into relation (1), we get the required interpolation polynomial. In practical applications, however, the solution of a system usually involves cumbersome calculations. We shall therefore seek the interpolation polynomial (1) in the form

$$\begin{aligned} P_n(x) = {} & a_0 (x - x_1)(x - x_2) \ldots (x - x_n) \\ & + a_1 (x - x_0)(x - x_2) \ldots (x - x_n) \\ & + a_2 (x - x_0)(x - x_1)(x - x_3) \ldots (x - x_n) + \\ & \ldots + a_n (x - x_0)(x - x_1)(x - x_2) \ldots (x - x_{n-1}). \end{aligned} \tag{4}$$

Setting $x = x_0$ in (4) and taking conditions (2) into account, we find that

$$y_0 = a_0 (x_0 - x_1)(x_0 - x_2) \ldots (x_0 - x_n),$$

whence

$$a_0 = \frac{y_0}{(x_0 - x_1)(x_0 - x_2) \ldots (x_0 - x_n)}.$$

Setting then $x = x_1$ in (4), we find that

$$y_1 = a_1 (x_1 - x_0)(x_1 - x_2) \ldots (x_1 - x_n),$$

whence

$$a_1 = \frac{y_1}{(x_1-x_0)(x_1-x_2)\ldots(x_1-x_n)}.$$

By analogy we find that

$$a_2 = \frac{y_2}{(x_2-x_0)(x_2-x_1)(x_2-x_3)\ldots(x_2-x_n)},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$a_n = \frac{y_n}{(x_n-x_0)(x_n-x_1)(x_n-x_2)\ldots(x_n-x_{n-1})}.$$

Substituting the values of the coefficients $a_0, a_1, \ldots, a_n$ obtained into formula (4), we get the required interpolation polynomial

$$P_n(x) = \frac{(x-x_1)(x-x_2)\ldots(x-x_n)}{(x_0-x_1)(x_0-x_2)\ldots(x_0-x_n)}\,y_0 + \frac{(x-x_0)(x-x_2)\ldots(x-x_n)}{(x_1-x_0)(x_1-x_2)\ldots(x_1-x_n)}\,y_1 + \ldots + \frac{(x-x_0)(x-x_1)\ldots(x-x_{n-1})}{(x_n-x_0)(x_n-x_1)\ldots(x_n-x_{n-1})}\,y_n. \tag{5}$$

Formula (5) is known as *Lagrange's formula for interpolation.*

**Example.** As a result of an experiment we have got, at the points $x_0 = 1$, $x_1 = 3$, $x_2 = 5$, the values of the function $f(x)$ equal to $y_0 = 2$, $y_1 = 1$, $y_2 = 8$ respectively. Find the second-degree polynomial $P_2(x)$ which approximates the function $f(x)$.

*Solution.* Using formula (5), we find that

$$P_2(x) = \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}\,y_0 + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}\,y_1 + \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}\,y_2 = \frac{(x-3)(x-5)}{(1-3)(1-5)}\,2 + \frac{(x-1)(x-5)}{(3-1)(3-5)}\,1 + \frac{(x-1)(x-3)}{(5-1)(5-3)}\,8,$$

$$P_2(x) = x^2 - \frac{9}{2}x + \frac{11}{2}.$$

**3. Newton's formula for interpolation.** We shall consider a special case when the difference $h$ between the neighbouring interpolation nodes is a constant quantity: $x_i - x_{i-1} = h$. We introduce the following designations:

$$\Delta y_0 = y_1 - y_0, \quad \Delta y_1 = y_2 - y_1, \quad \Delta y_2 = y_3 - y_2, \ldots,$$

$$\Delta^2 y_0 = \Delta y_1 - \Delta y_0, \quad \Delta^2 y_1 = \Delta y_2 - \Delta y_1, \quad \Delta^2 y_2 = \Delta y_3 - \Delta y_2, \ldots,$$

$$\Delta^3 y_0 = \Delta^2 y_1 - \Delta^2 y_0, \quad \Delta^3 y_1 = \Delta^2 y_2 - \Delta^2 y_1, \ldots,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\Delta^n y_0 = \Delta^{n-1} y_1 - \Delta^{n-1} y_0, \quad \Delta^n y_1 = \Delta^{n-1} y_2 - \Delta^{n-1} y_1, \ldots,$$

which are called *differences of the first, second, third, . . ., nth order.*

Let us find the interpolation polynomial of the $n$th degree which assumes the values $y_0, y_1, y_2, \ldots, y_n$ at the points $x_0, x_1 = x_0 + h, x_2 = x_0 + 2h, \ldots, x_n = x_0 + nh$ respectively. We shall first find the first-degree polynomial which assumes the values $y_0$ and $y_1$ at the points $x_0$ and $x_1 = x_0 + h$. Substituting the number $x_1 = x_0 + h$ for $x_1$ into formula (5), we get

$$P_1(x) = y_0 + \Delta y_0 \frac{x - x_0}{h}.$$

By analogy we find that

$$P_2(x) = y_0 + \Delta y_0 \frac{x - x_0}{h} + \frac{\Delta^2 y_0}{2!h^2}(x - x_0)(x - x_1),$$

$$P_3(x) = y_0 + \Delta y_0 \frac{x - x_0}{h} + \frac{\Delta^2 y_0}{2!h^2}(x - x_0)(x - x_1)$$
$$+ \frac{\Delta^3 y_0}{3!h^3}(x - x_0)(x - x_1)(x - x_2),$$

and, in general,

$$P_n(x) = y_0 + \Delta y_0 \frac{x - x_0}{h} + \frac{\Delta^2 y_0}{2!h^2}(x - x_0)(x - x_1)$$
$$+ \frac{\Delta^3 y_0}{3!h^3}(x - x_0)(x - x_1)(x - x_2) +$$
$$\ldots + \frac{\Delta^n y_0}{n!h^n}(x - x_0)(x - x_1)(x - x_2) \ldots (x - x_{n-1}). \tag{6}$$

Formula (6) defines the required polynomial and is known as *Newton's formula for interpolation.*

The interpolation problem has a unique solution and therefore Lagrange's and Newton's formulas for the given values of $x_i$ and $y_i$ are identical and differ only in the grouping of terms. Newton's formula is more convenient for applications since when new interpolation nodes are added, all the coefficients must be recalculated in Lagrange's formula, whereas only new terms are added and the old ones remain unchanged in Newton's formula.

There are other formulas for interpolation, *Hermite's formula* being most widely used. The problem is stated as follows: given $n$ nodes, $n$ values of the function $f(x)$ and $n$ values of its derivative $f'(x)$ at the nodes, we must find a polynomial of degree not higher than $2n - 1$ such that

$$P(x_i) = f(x_i) = y_i, \; P'(x_i) = f'(x_i) = y'_i,$$
$$i = 1, 2, \ldots, n.$$

We shall not present the solution of this problem here but only point out that if all $x_i$ are different, then there is a unique solution which can be found by analogy with the preceding one.

**4. The remainder of interpolation.** To find how close the interpolation polynomial $P_n(x)$ is to the function $f(x)$, it is necessary to investigate the difference between the function and the interpolation polynomial

$$R(x) = f(x) - P_n(x),$$

which is known as the *remainder of the interpolation.*

We assume that there is an $(n+1)$th continuous derivative $f^{(n+1)}(x)$ on the interval $[a, b]$. Then

$$R^{(n+1)}(x) = f^{(n+1)}(x), \quad a \leqslant x \leqslant b, \tag{7}$$

since $P_n^{(n+1)}(x) \equiv 0$. Assume that $x$ is any fixed number, $a \leqslant x \leqslant b$, which does not coincide with the interpolation nodes, and $t$ is a variable quantity, $a \leqslant t \leqslant b$. We set

$$\omega(x) = (x - x_0)(x - x_1)(x - x_2) \ldots (x - x_n)$$

and consider an auxiliary function

$$F(t) = R(t) - \frac{R(x)}{\omega(x)}\omega(t), \quad a \leqslant t \leqslant b,$$

on the interval $[a, b]$. The function $F(t)$ is evidently $n+1$ times differentiable on the interval $[a, b]$ and, by virtue of (7) and the fact that $\omega^{(n+1)}(t) = (n+1)!$, we have

$$F^{(n+1)}(t) = f^{(n+1)}(t) - (n+1)!\,\frac{R(x)}{\omega(x)}. \tag{8}$$

Furthermore, the function $F(t)$ vanishes at $n+2$ points: $x_0, x_1, \ldots, x_n$ and $x$ ($x$ is fixed). Therefore, by Rolle's theorem, its first derivative vanishes at least at $n+1$ points of the interval $[a, b]$, the second derivative vanishes at least at $n$ points of the interval, and so on. We find by induction that the $(n+1)$th derivative of the function $F(t)$ vanishes at least once within the interval $[a, b]$. Consequently, there is a point $\xi$, $a < \xi < b$, such that

$$F^{(n+1)}(\xi) = 0. \tag{9}$$

Setting $t = \xi$ in (8) and using (9), we find that

$$F^{(n+1)}(\xi) = f^{(n+1)}(\xi) - (n+1)!\,\frac{R(x)}{\omega(x)} = 0,$$

whence

$$R(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\omega(x), \quad a \leqslant x \leqslant b, \ a < \xi < b. \tag{10}$$

Relation (10) defines the remainder of interpolation. We designate the greatest value of the function $|f^{(n+1)}(x)|$ on the interval $[a, b]$ as $k$ and get a formula for evaluating the remainder for any $x \in [a, b]$:

$$|R(x)| \leqslant \frac{k}{(n+1)!}|\omega(x)|.$$

### 1.6.6. Methods of Approximation of the Roots of Equations

In this section we shall consider the problem of approximate calculation of the roots of the equation $f(x) = 0$, where $f(x)$ is a continuous function.

We know from elementary mathematics how to find the roots of the equation $f(x) = 0$ if $f(x)$ is a linear or quadratic function. When a function is more complicated, it is usually necessary to use various methods of approximation of the roots of equations. We shall discuss here the *"fork" method* and *Newton's method.*

**1. The "fork" method.** Assume that the root of the equation $f(x) = 0$ in question is an interior point of the interval $[a, b]$ and there are no other roots on $[a, b]$. Let the function $f(x)$ be continuous on $[a, b]$ and have values of different signs at the endpoints of the interval. In practical applications such an interval is usually found by an approximate estimation. We shall call any interval at whose endpoints $f(x)$ has values of different signs a "fork".

We assume, for definiteness, that $f(a) < 0$ and $f(b) > 0$. We bisect $[a, b]$ and choose that of the intervals obtained at whose endpoints $f(x)$ has different signs. We designate it as $[a_1, b_1]$. (If the value of $f(x)$ were zero at the midpoint of $[a, b]$, then the root would be found.) We bisect $[a_1, b_1]$ and again choose that of the intervals obtained at whose endpoints $f(x)$ has different signs, and so on. Continuing this process indefinitely, we get a sequence of nested intervals, "forks":

$$[a, b] \supset [a_1, b_1] \supset [a_2, b_2] \supset \ldots \supset [a_n, b_n] \supset \ldots,$$

which possess the property that $f(a_n) < 0$ and $f(b_n) > 0$ for any $n$. According to Theorem 2.13 on nested intervals, there is a point $c$ belonging to all the intervals, to which each of the sequences $\{a_n\}$ and $\{b_n\}$ converges.

We shall prove that the point $c$ is precisely the required root, i.e. $f(c) = 0$. Since $f(x)$ is continuous at the point $c$, each of the sequences $\{f(a_n)\}$ and $\{f(b_n)\}$ converges to $f(c)$. But then, in accordance with Theorem 2.10, we find from the conditions $f(a_n) < 0$ and $f(b_n) > 0$ that the inequalities $f(c) \leqslant 0$ and $f(c) \geqslant 0$ simultaneously hold true. Hence $f(c) = 0$, and this is what we wished to prove.

It is clear now how to approximate the root $x = c$. We can take the midpoint of the interval $[a_n, b_n]$, i.e. the point $(a_n + b_n)/2$, as the approximate value of this root. Since the length of $[a_n, b_n]$ is equal to $(b - a)/2^n$ and the distance from the root $c$ to the point $(a_n + b_n)/2$ does not exceed half the length of the interval $[a_n, b_n]$, the number $(a_n + b_n)/2$ differs from the exact value of the root by not more than $(b - a)/2^{n+1}$. Thus the method described makes it

possible to calculate the required root $c$ with any degree of accuracy if $n$ is sufficiently large. This method is very convenient since it requires uniform calculating operations. It is therefore often used when calculations are carried out by means of modern computers.

**2. Newton's method.** This is one of the most efficient methods of approximation of the roots of the equation $f(x) = 0$.

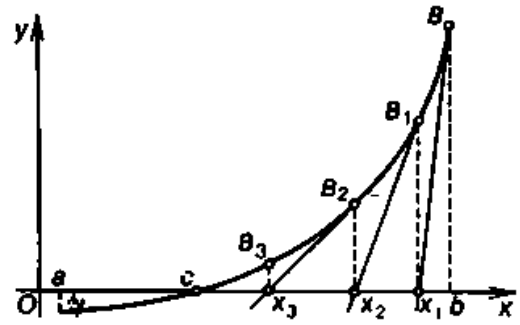


Fig. 92

Let, as before, $x = c$ be an interior point of $[a, b]$. We also assume that on $[a, b]$ the function $f(x)$ has derivatives $f'(x)$ and $f''(x)$ of constant sign and its values $f(a)$ and $f(b)$ are of different signs. Since the sign of $f'(x)$ is constant, the function $f(x)$ either increases or decreases on $[a, b]$ and, consequently, in both cases the graph of the function $y = f(x)$ cuts the $x$-axis only at one point, i.e. $x = c$ is the only root on $[a, b]$. Similarly, since the sign of $f''(x)$ is constant, the direction of convexity of the graph of the function $y = f(x)$ does not change on this interval.

For the sake of definiteness, we shall consider the case when $f'(x) > 0$ and $f''(x) > 0$. In this case $f(a) < 0$, $f(b) > 0$ and the graph is convex downwards (Fig. 92). We draw a tangent to the graph of the function $y = f(x)$ through the point $B(b, f(b))$. Its equation has the form

$$y - f(b) = f'(b)(x - b).$$

Setting $y = 0$, we find the abscissa of the point of intersection of the tangent with the $x$-axis:

$$x_1 = b - \frac{f(b)}{f'(b)}.$$

Since $\frac{f(b)}{f'(b)} > 0$, we have $x_1 < b$, and since the graph of the function $y = f(x)$ lies not lower than the tangent, we have $x_1 > c$. Thus $c < x_1 < b$. We take the point $x_1$ as the first approximation of the value of the root. Then we draw a tangent to the graph through the point $B_1(x_1, f(x_1))$ and by analogy take the point $x_2$ as the second approximation of the value of the root:

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$$

In this case $c < x_2 < x_1$. Continuing this process indefinitely, we get the following formula for any $n$:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \tag{1}$$

which expresses $x_{n+1}$ in terms of $x_n$. Thus we have a sequence of approximate values of the root $c$, with

$$x_1 > x_2 > x_3 > \ldots > x_n > x_{n+1} > \ldots > c. \qquad (2)$$

Formula (1) is the fundamental formula of Newton's method, i.e. the method of successive approximations (iterations) which is constructed by means of formula (1).

We shall prove that the sequence $\{x_n\}$ converges to the required root $c$ and estimate the error, i.e. the deviation of the approximate value $x_n$ from the exact value of the root $c$. Indeed, by virtue of (2), the sequence $\{x_n\}$ decreases and is bounded from below by the number $c$. Consequently, in accordance with Theorem 2.12, it has a limit $c' \geqslant c$. Passing to the limit in formula (1) and taking into account the continuity of $f(x)$ and $f'(x)$, we obtain

$$c' = c' - \frac{f(c')}{f'(c')},$$

whence it follows that $f(c') = 0$, i.e. $c'$ is a root of the equation $f(x) = 0$. But since there is only one root, $c$, on $[a, b]$, it follows that $c' = c$. Thus the sequence $\{x_n\}$ converges to the root $c$.

Let us now estimate the deviation of the $n$th approximation $x_n$ from the exact value of the root $c$. Applying Lagrange's formula to the expression $f(x_n) = f(x_n) - f(c)$, we have $f(x_n) = (x_n - c) f'(\xi_n)$, where $c < \xi_n < x_n$. From this we obtain the following estimate:

$$|x_n - c| \leqslant \frac{|f(x_n)|}{m}, \qquad (3)$$

where $m$ is the least value of $|f'(x)|$ on the interval $[a, b]$. Formula (3) makes it possible to estimate the deviation of the approximate value $x_n$ from the exact value of the root $c$ in terms of the value of the modulus of the function $f(x)$ at the point $x_n$. Note that the estimate (3) is valid not only for Newton's method but, in general, for any method of approximation of a root provided that $m \neq 0$.

We have considered the case when $f'(x) > 0$ and $f''(x) > 0$ on $[a, b]$. Depending on the combination of signs of $f'(x)$ and $f''(x)$, three more cases are possible: (1) $f'(x) > 0$, $f''(x) < 0$; (2) $f'(x) < 0$, $f''(x) > 0$; (3) $f'(x) < 0$, $f''(x) < 0$, in each of which the substantiation of Newton's method is similar to the case we have considered.

**Example.** Calculate the root of the equation $x^2 - 5 = 0$ using Newton's method.

*Solution.* We consider the function $f(x) = x^2 - 5$. This function is continuous throughout the number line. We find the interval at whose endpoints the function $f(x)$ has values of different signs. Since $f(2) = -1$ and $f(3) = 4$, the closed interval $[2, 3]$ is of this kind. It includes the required root of the equation. The function

$f(x)$ has continuous positive derivatives $f'(x) = 2x$ and $f''(x) = 2$ on this interval. Consequently, we must draw the first tangent to the graph of the function $y = f(x)$ through the point (3, 4). We set $x_0 = 3$ in formula (1) and get the first approximation of the root: $x_1 = 3 - \frac{4}{2\cdot 3} = 2\frac{1}{3}$. Now we set $x_1 = 2\frac{1}{3}$ in formula (1) and get the second approximation of the root: $x_2 = 2\frac{1}{3} - \frac{2}{21} = 2\frac{5}{21}$, and, finally, setting $x_2 = 2\frac{5}{21}$ in formula (1), we get the third approximation of the root: $x_3 = 2\frac{2}{21} - \frac{2}{987} = 2.23607$ and so on.

To find the error of the approximation $x_3$, we use formula (3). Since the derivative $f'(x) = 2x$ increases on [2, 3], the value $f'(2) = 4$ is its least value on this interval, i.e. $m = 4$. Let us find $f(x_3)$: $f(x_3) = x_3^2 - 5 = (2.23607)^2 - 5 = 0.00001$. Using now formula (3), we have

$$|x_3 - c| < \frac{0.00001}{4} = 0.0000025 = 2.5\cdot 10^{-6}.$$

**If, by the hypothesis, such an accuracy of calculation of the root is sufficient, then the process of constructing approximations should be terminated, if not, the process must be continued.**

Chapter 7

# THE INDEFINITE INTEGRAL

## 1.7.1. The Antiderivative and the Indefinite Integral

**1. The concept of an antiderivative of a function.** Seeking a derivative of a given function is one of the most important problems of differential calculus. Various problems of analysis and its numerous applications in geometry, mechanics, physics and engineering lead to the converse problem: using the given function $f(x)$, we must find a function $F(x)$ whose derivative would be equal to the function $f(x)$, i.e. $F'(x) = f(x)$.

Finding a function from the known derivative of that function is one of the fundamental problems of integral calculus.

**Definition 1.** *The function $F(x)$ is an antiderivative of the function $f(x)$ on an interval $X$ if for all values of $x$ from this interval the equality $F'(x) = f(x)$ holds true.*

We shall consider some examples.

1. The function $F(x) = \sin x$ is an antiderivative of the function $f(x) = \cos x$ throughout the number line since $(\sin x)' = \cos x$ for any value of $x$.

2. The function $F(x) = x^3$ is an antiderivative of the function $f(x) = 3x^2$ throughout the number line since $(x^3)' = 3x^2$ at every point $x$.

3. The function $F(x) = \sqrt{1 - x^2}$ is an antiderivative of the function $f(x) = -x/\sqrt{1 - x^2}$ on the interval $(-1, +1)$ since $(\sqrt{1 - x^2})' = -x/\sqrt{1 - x^2}$ at any point $x$ of this interval.

The problem of finding an antiderivative, being given a function $f(x)$, does not have a unique solution. Indeed, if $F(x)$ is an antiderivative of $f(x)$, i.e. $F'(x) = f(x)$, then the function $F(x) + C$, where $C$ is an arbitrary constant, is also an antiderivative of $f(x)$ since $[F(x) + C]' = f(x)$ for any number $C$. For instance, $f(x) = \cos x$ has not only an antiderivative $\sin x$ but also an antiderivative $\sin x + C$ since $(\sin x + C)' = \cos x$.

We shall now show that the set of functions $F(x) + C$, where $F(x)$ is an antiderivative of the function $f(x)$ and $C$ is an arbitrary constant, exhausts all the antiderivatives of the function $f(x)$.

**Lemma 7.1.** *A function whose derivative on an interval $X$ is zero is constant on this interval.*

*Proof.* Let the derivative of the function $f(x)$ be zero at all points of the interval $X$, i.e. $f'(x) = 0$. By Lagrange's theorem, for any two points $x_1, x_2 \in X$ we obtain

$$f(x_2) - f(x_1) = f'(\xi)(x_2 - x_1), \quad x_1 < \xi < x_2.$$

Since $f'(\xi) = 0$, it follows that $f(x_2) = f(x_1)$. And this means that the values of the function at all points of the interval are the same, i.e. $f(x) = C$, where $C$ is a number. ■

**Theorem 7.1.** *If $F(x)$ is an antiderivative of the function $f(x)$ on an interval $X$, then any other antiderivative of $f(x)$ on this interval can be represented in the form $F(x) + C$, where $C$ is an arbitrary constant.*

*Proof.* Let $\Phi(x)$ be any other antiderivative of the function $f(x)$ on the interval $X$, i.e. $\Phi'(x) = f(x)$. Then, for any $x \in X$, we have

$$[\Phi(x) - F(x)]' = \Phi'(x) - F'(x) = f(x) - f(x) = 0,$$

and (according to Lemma 7.1) this means that the function $\Phi(x) - F(x)$ is constant, i.e. $\Phi(x) - F(x) = C$, where $C$ is a number. Consequently, $\Phi(x) = F(x) + C$. ■

It follows from the theorem we have proved that the set of functions $F(x) + C$, where $F(x)$ is one of the antiderivatives of the function $f(x)$ and $C$ is an arbitrary constant, exhausts the whole family of antiderivatives of $f(x)$.

**2. The indefinite integral.**

**Definition 2.** *If the function $F(x)$ is an antiderivative of the function $f(x)$ on an interval $X$, then the set of functions $F(x)+C$, where $C$ is an arbitrary constant, is an indefinite integral of the function $f(x)$ on this interval and is designated as*

$$\int f(x)\,dx = F(x) + C.$$

In this case the function $f(x)$ is the *integrand*, $f(x)\,dx$ is the *element of integration* and the variable $x$ is the *integration variable*.

Thus the symbol $\int f(x)\,dx$ denotes the collection of all antiderivatives of the function $f(x)$. But sometimes we shall understand it to be any element of that collection, i.e. any one of the antiderivatives.

Finding a function being given its derivative, or, what is the same, seeking an indefinite integral being given an integrand, is the *integration* of that function. Integration is the inverse of differentiation. To verify whether integration was carried out correctly, it is sufficient to differentiate the result and obtain the integrand.

In this chapter we do not consider the problem of existence of antiderivatives (and, hence, indefinite integrals) for broad classes of functions. In 1.8.6 we shall prove that any function continuous on a closed interval has an antiderivative on that interval (and, consequently, an indefinite integral).

**Examples.**

1. $\int 3x^2\,dx = x^3 + C$ since $(x^3+C)' = 3x^2$.
2. $\int \cos x\,dx = \sin x + C$ since $(\sin x + C)' = \cos x$.
3. $\int \frac{1}{x}\,dx = \ln|x| + C$ since $(\ln|x| + C)' = \frac{1}{x}$.
4. $\int e^{-2x}\,dx = -\frac{1}{2}e^{-2x} + C$ since $\left(-\frac{1}{2}e^{-2x} + C\right)' = e^{-2x}$ etc.

### 1.7.2. Principal Properties of the Indefinite Integral

The definition of the indefinite integral yields the following properties.

1°. *The derivative of an indefinite integral is equal to the integrand, the differential of an indefinite integral is equal to the element of integration, i.e.*

$$\left(\int f(x)\,dx\right)' = f(x) \text{ and } d\int f(x)\,dx = f(x)\,dx.$$

Indeed,

$$\left(\int f(x)\,dx\right)' = (F(x)+C)' = F'(x) = f(x)$$

and

$$d\int f(x)\,dx = \left(\int f(x)dx\right)'\,dx = f(x)\,dx.$$

2°. *An indefinite integral of the differential of a function is equal to the sum of that function and the arbitrary constant, i.e.*

$$\int dF(x) = F(x) + C.$$

Indeed, since $dF(x) = F'(x)\,dx$, it follows that

$$\int F'(x)\,dx = F(x) + C.$$

3°. *We can put the constant factor before the sign of the integral, i.e. if* $k = \text{const} \neq 0$, *then*

$$\int kf(x)\,dx = k\int f(x)\,dx.$$

Indeed, let $F(x)$ be an antiderivative of the function $f(x)$, i.e. $F'(x) = f(x)$. Then $kF(x)$ is an antiderivative of the function $kf(x)$: $(kF(x))' = kF'(x) = kf(x)$. From this it follows that

$$k\int f(x)\,dx = k[F(x)+C] = kF(x) + C_1 = \int kf(x)\,dx,$$

where $C_1 = kC$.

4°. *An indefinite integral of the algebraic sum of two functions is equal to the algebraic sum of the integrals of those functions taken separately, i.e.*

$$\int [f(x) \pm g(x)]\,dx = \int f(x)\,dx \pm \int g(x)\,dx.$$

Indeed, let $F(x)$ and $G(x)$ be antiderivatives of the functions $f(x)$ and $g(x)$:

$$F'(x) = f(x), \quad G'(x) = g(x).$$

Then the functions $F(x) \pm G(x)$ are antiderivatives of the functions $f(x) \pm g(x)$. Consequently,

$$\int f(x)\,dx \pm \int g(x)\,dx = [F(x)+C_1] \pm [G(x)+C_2]$$

$$= [F(x) \pm G(x)] + (C_1 \pm C_2) = [F(x) \pm G(x)] + C = \int [f(x) \pm g(x)]\,dx.$$

Note that this property is valid for any finite number of functions being added.

### 1.7.3. Table of Fundamental Integrals

Here is a *table of fundamental integrals*. A part of the formulas presented in the table follow directly from the definition of integration as the inverse of differentiation and the table of derivatives, and it is easy to verify the validity of the other formulas by differentiation.

I. $\int x^{\alpha}\,dx = \frac{x^{\alpha+1}}{\alpha+1} + C \quad (\alpha \neq -1).$

II. $\int \frac{dx}{x} = \ln|x| + C.$

III. $\int \frac{dx}{1+x^2} = \arctan x + C.$

IV. $\int \frac{dx}{\sqrt{1-x^2}} = \arcsin x + C.$

V. $\int a^x\,dx = \frac{a^x}{\ln a} + C \quad (0 < a \neq 1).$

VI. $\int e^x\,dx = e^x + C.$

VII. $\int \sin x\,dx = -\cos x + C.$

VIII. $\int \cos x\,dx = \sin x + C.$

IX. $\int \frac{dx}{\cos^2 x} = \tan x + C.$

X. $\int \frac{dx}{\sin^2 x} = -\cot x + C.$

XI. $\int \frac{dx}{x^2-a^2} = \frac{1}{2a}\ln\left|\frac{x-a}{x+a}\right| + C \quad (a \neq 0).$

XII. $\int \frac{dx}{\sqrt{x^2+k}} = \ln\left|x + \sqrt{x^2+k}\right| + C.$

XIII. $\int \frac{dx}{x^2+a^2} = \frac{1}{a}\arctan\frac{x}{a} + C.$

XIV. $\int \frac{dx}{\sqrt{a^2-x^2}} = \arcsin\frac{x}{a} + C.$

The integrals given in this table are usually called *tabular integrals*.

### 1.7.4. Principal Methods of Integration

**1. Direct integration.** Calculation of integrals by a direct use of the table of the simplest integrals and the principal properties of indefinite integrals is called a *direct integration*.

**Examples.**

1. $\int \left(5\cos x+2-3x^2+\frac{1}{x}-\frac{4}{x^2+1}\right) dx=5\int \cos x\, dx+2\int dx-3\int x^2\, dx+\int \frac{dx}{x}-4\int \frac{dx}{x^2+1}=5\sin x+2x-x^3+\ln|x|-4\arctan x+C.$

2. $\int \left(\sin\frac{x}{2}+\cos\frac{x}{2}\right)^2 dx=\int \left(\sin^2\frac{x}{2}+2\sin\frac{x}{2}\cos\frac{x}{2}+\cos^2\frac{x}{2}\right) dx=\int (1+\sin x)\, dx=\int dx+\int \sin x\, dx=x-\cos x+C.$

3. $\int \tan^2 x\, dx=\int (\sec^2 x-1)\, dx=\int \frac{dx}{\cos^2 x}-\int dx=\tan x-x+C.$

**2. The method of substitution.** In many cases the introduction of a new integration variable makes it possible to reduce seeking a given integral to seeking a tabular integral, i.e. to pass to a direct integration. This method is known as the *method of substitution*, or the *method of a change of variable*. It is based on the following theorem.

**Theorem 7.2.** *Assume that the function $x=\varphi(t)$ is defined and differentiable on an interval $T$ and let $X$ be the set of values of this function on which the function $f(x)$ is defined. Then, if the function $f(x)$ has an antiderivative on the set $X$, then the formula*

$$\int f(x)\, dx\Big|_{x=\varphi(t)}=\int f[\varphi(t)]\,\varphi'(t)\, dt \tag{1}$$

*holds true on the set $T$.*

*Proof.* Let $F(x)$ be an antiderivative of $f(x)$ on the set $X$. Let us consider a composite function $F[\varphi(t)]$ on the set $T$. By the rule of differentiation of a composite function, we take into account that $F'(x)=f(x)$ and obtain

$$(F[\varphi(t)])'=F'_x[\varphi(t)]\,\varphi'(t)=f[\varphi(t)]\,\varphi'(t),$$

i.e. the function $f[\varphi(t)]\,\varphi'(t)$ has an antiderivative $F[\varphi(t)]$ on the set $T$ and, consequently,

$$\int f[\varphi(t)]\,\varphi'(t)\, dt=F[\varphi(t)]+C.$$

Noting that $F[\varphi(t)]+C=(F(x)+C)\Big|_{x=\varphi(t)}=\int f(x)\, dx\Big|_{x=\varphi(t)}$, we obtain formula (1). ■

Formula (1) is known as the *formula for a change of variable in an indefinite integral.*

**Example 1.** Calculate the integral $\int \frac{x^3}{(x-1)^2}\, dx$.

*Solution.* We set $x-1=t$ and then $x=t+1$. Hence $dx=dt$. From formula (1) we have

$$\int \frac{x^3}{(x-1)^2}\,dx=\int \frac{(t+1)^3}{t^2}\,dt=\int \left(t+3+\frac{3}{t}+\frac{1}{t^2}\right)dt$$
$$=\frac{1}{2}t^2+3t+3\ln|t|-\frac{1}{t}+C.$$

Returning to the variable $x$, we finally get

$$\int \frac{x^3}{(x-1)^2}\,dx=\frac{1}{2}(x-1)^2+3(x-1)+3\ln|x-1|-\frac{1}{x-1}+C.$$

**Remark.** When a change of variable is performed in an indefinite integral, it is sometimes more convenient to define $t$ as a function of $x$ rather than $x$ as a function of $t$.

**Example 2.** Calculate the integral $\int \frac{x^4\,dx}{x^5+7}$.

*Solution.* We set $x^5+7=t$, $dt=5x^4\,dx$. Then

$$\int \frac{x^4\,dx}{x^5+7}=\frac{1}{5}\int \frac{dt}{t}=\frac{1}{5}\ln|t|+C$$

so that

$$\int \frac{x^4\,dx}{x^5+7}=\frac{1}{5}\ln|x^5+7|+C.$$

Note that it is not always easy to choose a successful substitution. To overcome this difficulty, it is necessary to master the technique of differentiation and know the tabular integrals very well.

**Example 3.** Calculate the integral $\int \frac{dx}{\sqrt{x^2+a}}$.

*Solution.* We set $\sqrt{x^2+a}+x=t$, whence $\left(\frac{x}{\sqrt{x^2+a}}+1\right)dx=dt$. Thus

$$dx=\frac{\sqrt{x^2+a}}{\sqrt{x^2+a}+x}\,dt$$

so that

$$\int \frac{dx}{\sqrt{x^2+a}}=\int \frac{dt}{t}=\ln|t|+C=\ln|\sqrt{x^2+a}+x|+C.$$

**Example 4.** Calculate the integral $\int \sin^n x\cos x\,dx$.

*Solution.* We set $t=\sin x$, $dt=\cos x\,dx$. Then

$$\int \sin^n x\cos x\,dx=\int t^n\,dt=\begin{cases}\frac{t^{n+1}}{n+1}+C=\frac{\sin^{n+1}x}{n+1}+C & \text{for } n\neq-1,\\ \ln|t|+C=\ln|\sin x|+C & \text{for } n=-1.\end{cases}$$

**Example 5.** Calculate the integral $\int \frac{x\,dx}{(x^2+1)^n}$, $n \neq 1$.

*Solution.* We set $x^2+1=t$, $2x\,dx=dt$. Then

$$\int \frac{x\,dx}{(x^2+1)^n} = \frac{1}{2}\int \frac{dt}{t^n} = -\frac{1}{2(n-1)}\,\frac{1}{t^{n-1}}+C$$
$$= -\frac{1}{2(n-1)}\,\frac{1}{(x^2+1)^{n-1}}+C.$$

For $n=1$ we find by analogy that

$$\int \frac{x\,dx}{x^2+1} = \frac{1}{2}\ln(x^2+1)+C.$$

**3. The method of integration by parts.** This method is based on the use of the formula for the differentiation of the product of two functions.

**Theorem 7.3.** *Assume that the functions $u(x)$ and $v(x)$ are defined and differentiable on an interval $X$ and let the function $u'(x)\,v(x)$ have an antiderivative on this interval. Then, on the interval $X$, the function $u(x)\,v'(x)$ also has an antiderivative and there holds a formula*

$$\int u(x)\,v'(x)\,dx = u(x)\,v(x) - \int v(x)\,u'(x)\,dx. \tag{2}$$

*Proof.* It follows from the relation

$$[u(x)\,v(x)]' = u'(x)\,v(x) + u(x)\,v'(x)$$

that

$$u(x)\,v'(x) = [u(x)\,v(x)]' - u'(x)\,v(x).$$

The function $u(x)\,v(x)$ is an antiderivative of the function $[u(x)\,v(x)]'$ on the interval $X$. The function $u'(x)\,v(x)$ has an antiderivative on $X$ by the hypothesis of the theorem. Consequently, the function $u(x)\,v'(x)$ also has an antiderivative on the interval $X$. Integrating the last relation, we get formula (2). ■

Formula (2) is known as the *formula for integration by parts in an indefinite integral.*

Since $v'(x)\,dx = dv$ and $u'(x)\,dx = du$, it can be rewritten as

$$\int u\,dv = u\cdot v - \int v\,du.$$

This formula makes it possible to reduce the calculation of $\int u\,dv$ to the calculation of the integral $\int v\,du$, which may prove to be simpler.

**Examples.**

1. $\int \underbrace{\arctan x}_{u}\, \underbrace{dx}_{dv} = \left| \begin{array}{l} u = \arctan x, \ du = \frac{dx}{1+x^2} \\ dv = dx, \ \int dv = \int dx, \ v = x \end{array} \right|^{*} =$

$\underbrace{x}_{v} \underbrace{\arctan x}_{u} - \int \underbrace{x}_{v} \underbrace{\frac{dx}{1+x^2}}_{du} = x \arctan x - \frac{1}{2} \int \frac{d(x^2+1)}{1+x^2} = x \arctan x -$

$\frac{1}{2} \ln (1 + x^2) + C.$

2. $\int x e^x \, dx = \left| \begin{array}{l} u = x, \ du = dx \\ dv = e^x \, dx, \ \int dv = \int e^x \, dx, \ v = e^x \end{array} \right| =$

$\int \underbrace{x}_{u} \underbrace{de^x}_{dv} = \underbrace{x}_{u} \underbrace{e^x}_{v} - \int \underbrace{e^x}_{v} \underbrace{dx}_{du} = xe^x - e^x + C.$

3. $\int x \ln x \, dx = \left| \begin{array}{l} u = \ln x, \ du = \frac{dx}{x} \\ dv = x \, dx, \ \int dv = \int x \, dx, \ v = \frac{x^2}{2} \end{array} \right| =$

$\int \ln x d \left( \frac{x^2}{2} \right) = \frac{x^2}{2} \ln x - \int \frac{x^2}{2} \frac{1}{x} \, dx = \frac{x^2}{2} \ln x - \frac{x^2}{4} + C.$

Sometimes, to calculate an integral, the formula for integration by parts must be used several times.

4. $\int x^2 \cos x \, dx = \left| \begin{array}{l} u = x^2, \ du = 2x \, dx \\ dv = \cos x \, dx, \ \int dv = \int \cos x \, dx, \ v = \sin x \end{array} \right| =$

$\int x^2 d (\sin x) = x^2 \sin x - 2 \int x \sin x \, dx =$

$\left| \begin{array}{l} u = x, \ du = dx \\ dv = \sin x \, dx, \ \int dv = \int \sin x \, dx, \ v = -\cos x \end{array} \right| = x^2 \sin x +$

$2 \int x d \cos x = x^2 \sin x + 2x \cos x - 2 \int \cos x \, dx = x^2 \sin x + 2x \cos x -$

$2 \sin x + C.$

Thus the integral $\int x^2 \cos x \, dx$ has been calculated by twice using integration by parts.

---

* Vertical lines separate here auxiliary notations. It must also be pointed out that we can take any function of the form $x + C$ as $v$, where $C$ is a constant. We have taken $v = x$, i.e. $C = 0$.

In conclusion, we shall calculate the integral

$$I_n = \int \frac{dx}{(x^2+1)^n}$$

($n$ is a positive integer), which we shall need in the following section. For $n = 1$ we have a tabular integral

$$I_1 = \arctan x + C.$$

Let $n > 1$. Representing unity in the numerator as the difference $(x^2 + 1) - x^2$, we obtain

$$I_n = \int \frac{dx}{(x^2+1)^{n-1}} - \int \frac{x^2\,dx}{(x^2+1)^n}.$$

In the second integral we apply the method of integration by parts:

$$u = x,\ du = dx,\ dv = \frac{x\,dx}{(x^2+1)^n},\ v = \int \frac{x\,dx}{(x^2+1)^n} = -\frac{1}{(2n-2)(x^2+1)^{n-1}}$$

(see Item 2, Example 5), and then

$$\int \frac{x^2\,dx}{(x^2+1)^n} = -\frac{x}{(2n-2)(x^2+1)^{n-1}} + \int \frac{dx}{(2n-2)(x^2+1)^{n-1}}.$$

Consequently,

$$I_n = I_{n-1} + \frac{x}{(2n-2)(x^2+1)^{n-1}} - \frac{1}{2n-2} I_{n-1},$$

whence

$$I_n = \frac{x}{(2n-2)(x^2+1)^{n-1}} + \frac{2n-3}{2n-2} I_{n-1}.$$

We have thus expressed the integral $I_n$ in terms of $I_{n-1}$:

$$\int \frac{dx}{(x^2+1)^n} = \frac{x}{(2n-2)(x^2+1)^{n-1}} + \frac{2n-3}{2n-2} \int \frac{dx}{(x^2+1)^{n-1}} \quad (n > 1). \qquad (3)$$

Formulas of type (3) are called the *recursion formulas*.

**Example 5.** Calculate $\int \frac{dx}{(x^2+1)^3}$.

*Solution.* From the recursion formula (3) we have

$$\int \frac{dx}{(x^2+1)^3} = \frac{x}{4(x^2+1)^2} + \frac{3}{4} \int \frac{dx}{(x^2+1)^2},$$

$$\int \frac{dx}{(x^2+1)^2} = \frac{x}{2(x^2+1)} + \frac{1}{2} \int \frac{dx}{x^2+1},$$

and

$$I_1 = \int \frac{dx}{x^2+1} = \arctan x,$$

and therefore the final result is

$$\int \frac{dx}{(x^2+1)^3} = \frac{x}{4(x^2+1)^2} + \frac{3x}{8(x^2+1)} + \frac{3}{8}\arctan x + C.$$

### 1.7.5. Integration of Rational Functions

*Rational functions*, i.e. functions which can be represented as a fraction

$$\frac{P(x)}{Q(x)},$$

where $P(x)$ and $Q(x)$ are polynomials, form a significant class of functions whose integrals can be always expressed in terms of elementary functions.

If the degree of a polynomial in the numerator is equal to or greater than the degree of the polynomial in the denominator, we can perform division and obtain

$$\frac{P(x)}{Q(x)} = W(x) + \frac{R(x)}{Q(x)}, \tag{1}$$

where $W(x)$ is a polynomial and $R(x)$ is a polynomial of a degree lower than that of $Q(x)$.

**Examples.**

1. $\dfrac{x^6+x^3-x^2+1}{x^3-2x+1} = x^2+3-\dfrac{2x^2-6x+2}{x^3-2x+1}$.

2. $\dfrac{x^3+x+1}{x^2+1} = x+\dfrac{1}{x^2+1}$.

It is proved in higher algebra that every polynomial can be represented as a product

$$Q(x) = A(x-\alpha)(x-\beta)\ldots(x-\gamma),$$

where $A$ is the coefficient in the highest degree of the polynomial $Q(x)$ and $\alpha, \beta, \ldots, \gamma$ are the roots of the equation $Q(x) = 0$. The factors $(x-\alpha)(x-\beta)\ldots(x-\gamma)$ are *elementary factors*. If there are coincident factors among them, then we can group them and get a representation

$$Q(x) = A(x-\alpha)^r(x-\beta)^s\ldots(x-\gamma)^t, \tag{2}$$

where $r, s, \ldots, t$ are integers which are the *multiplicities of the roots* $\alpha, \beta, \ldots, \gamma$ respectively, with $r+s+\ldots+t = n$, $n$ being the degree of the polynomial $Q(x)$.

There may be complex roots among the roots of representation (2). It is proved in algebra that if $\alpha = a+bi$ is an $r$-tuple complex root of the polynomial with real coefficients, then the polynomial also has an $r$-tuple root $\bar{\alpha} = a-bi$ conjugate to $\alpha$. In other words, if

representation (2) includes a factor $(x-\alpha)^r$, where $\alpha = a + bi$ $(b \neq 0)$, then it also includes a factor $(x-\bar{\alpha})^r$. Multiplying these two factors, we obtain

$$\begin{aligned}(x-\alpha)^r(x-\bar{\alpha})^r &= \{[x-(a+bi)]\,[(x-(a-bi)]\}^r \\ &= [x^2 - x(a+bi) - x(a-bi) + a^2 + b^2]^r \\ &= (x^2 - 2ax + a^2 + b^2)^r = (x^2 + 2px + q)^r,\end{aligned}$$

where $p = -a$, $q = a^2 + b^2$, $p^2 - q < 0$, $p$ and $q$ being real numbers.

Doing the same to the other complex roots, we write representation (2) in the form

$$Q(x) = A(x-\alpha)^r(x-\beta)^s \ldots (x^2+2px+q)^t(x^2+2ux+v)^\eta \ldots, \tag{3}$$

where $\alpha, \beta, \ldots, p, q, u, v, \ldots$ are real numbers.

The following theorem is proved in higher algebra.

**Theorem 7.4.** *If the degree of the polynomial in the numerator of the rational function* $\frac{R(x)}{Q(x)}$ *is lower than the degree of the polynomial in the denominator and the polynomial* $Q(x)$ *is in form* (3). *then the function can be uniquely represented in the form*

$$\frac{R(x)}{Q(x)} = \frac{A_1}{(x-\alpha)} + \frac{A_2}{(x-\alpha)^2} + \ldots + \frac{A_r}{(x-\alpha)^r} +$$

$$\ldots + \frac{M_1x+N_1}{x^2+2px+q} + \frac{M_2x+N_2}{(x^2+2px+q)^2} + \ldots + \frac{M_tx+N_t}{(x^2+2px+q)^t} + \ldots, \tag{4}$$

*where* $A_1, A_2, \ldots, A_r, \ldots, M_1, N_1, M_2, N_2, \ldots, M_t, N_t, \ldots$ *are real numbers.*

Expression (4) is known as the *expansion of a rational function into partial fractions.*

Relation (4) holds true for all $x$ which are not real roots of the polynomial $Q(x)$.

To find the numbers $A_1, A_2, \ldots, A_r, \ldots, M_1, N_1, \ldots, M_t, N_t, \ldots$, we multiply both sides of expansion (4) with the still unknown $A_1, A_2, \ldots$ by $Q(x)$. Since the equality of the polynomial $R(x)$ and the polynomial which results on the right-hand side must hold true for all $x$, the coefficients in the equal powers of $x$ must be equal. Equating them to one another, we get a system of first-degree equations from which we can find the unknown numbers $A_1, A_2, \ldots, A_r, \ldots, M_1, N_1, \ldots, M_t, N_t, \ldots$.

This method of seeking the coefficients of the expansion of rational functions is known as the *method of undetermined coefficients.*

**Example 1.** Expand the rational function $\frac{2x-1}{x^2-5x+6}$ into partial fractions.

*Solution.* Since $x^2 - 5x + 6 = (x - 3)(x - 2)$, we have, from relation (4),

$$\frac{2x-1}{x^2-5x+6} = \frac{A}{x-3} + \frac{B}{x-2}.$$

Multiplying both sides of the relation by $x^2 - 5x + 6$, we get

$$2x - 1 = A(x - 2) + B(x - 3),$$

or

$$2x - 1 = (A + B)x - 2A - 3B.$$

Equating the coefficients in the equal powers of $x$ to one another, we get a system of first-degree equations with respect to $A$ and $B$:

$$\begin{cases} A + B = 2, \\ 2A + 3B = 1, \end{cases}$$

whence $A = 5$ and $B = -3$.

Thus

$$\frac{2x-1}{x^2-5x+6} = \frac{5}{x-3} - \frac{3}{x-2}.$$

**Example 2.** Find the expansion of the rational function $\frac{x^2-1}{x(x^2+1)^2}$ into partial fractions.

*Solution.* The quadratic trinomial $x^2 + 1$ has complex roots and therefore, from formula (4), we get

$$\frac{x^2-1}{x(x^2+1)^2} = \frac{A}{x} + \frac{Bx+C}{x^2+1} + \frac{Dx+E}{(x^2+1)^2}.$$

Multiplying both sides of the relation by $x(x^2 + 1)^2$, we obtain

$$x^2 - 1 = A(x^2 + 1)^2 + (Bx + C)(x^2 + 1)x + (Dx + E)x,$$

or

$$x^2 - 1 = (A + B)x^4 + Cx^3 + (2A + B + D)x^2 + (C + E)x + A.$$

Equating the coefficients in $x^0$, $x^1$, $x^2$, $x^3$, $x^4$ to one another, we arrive at a system of equations

$$\begin{cases} x^4: & A + B = 0, \\ x^3: & C = 0, \\ x^2: & 2A + B + D = 0, \\ x^1: & C + E = 0, \\ x^0: & A = -1, \end{cases}$$

solving which, we find that $A = -1$, $B = 1$, $C = 0$, $D = 2$, $E = 0$, and therefore the required expansion has the form

$$\frac{x^2-1}{x(x^2+1)^2} = -\frac{1}{x} + \frac{x}{x^2+1} + \frac{2x}{(x^2+1)^2}.$$

It follows from the aforesaid that the problem of integration of the rational function (1) reduces to the integration of the polynomial $W(x) = a_0x^m + a_1x^{m-1} + \ldots + a_m$ whose integral is tabular, i.e.

$$\int W(x)\,dx = a_0\frac{x^{m+1}}{m+1} + a_1\frac{x^m}{m} + \ldots + a_mx + C,$$

and to the integration of the rational function $\frac{R(x)}{Q(x)}$, and this leads to finding integrals of the following four types*

I. $\int \frac{A}{x-\alpha}\,dx = A\ln|x-\alpha| + C;$

II. $\int \frac{A}{(x-\alpha)^r}\,dx = -\frac{A}{(r-1)(x-\alpha)^{r-1}} + C \quad (r > 1);$

III. $\int \frac{Ax+B}{x^2+2px+q}\,dx;$

IV. $\int \frac{Ax+B}{(x^2+2px+q)^r}\,dx \quad (r > 1).$

The polynomial $x^2 + 2px + q$ does not have real roots so that $p^2 - q < 0$.

We shall calculate the integral of type III which is often encountered in applications. We isolate a perfect square from the trinomial in the denominator:

$$x^2 + 2px + q = (x + p)^2 + q - p^2.$$

This representation prompts the substitution $x + p = t$, whence we get $x = t - p$, $dx = dt$. Then we set $q - p^2 = h > 0$ and pass to the variable $t$. As a result, the integral is reduced to the form

$$\int \frac{Ax+B}{x^2+2px+q}\,dx = \int \frac{At+B-Ap}{t^2+h}\,dt$$
$$= \frac{1}{2}A\int \frac{2t\,dt}{t^2+h} + (B - A\cdot p)\int \frac{dt}{t^2+h}.$$

The first integral on the right-hand side is taken directly

$$\int \frac{2t\,dt}{t^2+h} = \ln|t^2+h| + C = \ln|x^2+2px+q| + C.$$

The second integral can be calculated by formula XIII of the table of fundamental integrals.

**Example 3.** Calculate $\int \frac{6x+5}{x^2+4x+9}\,dx$.

*Solution.* We isolate a perfect square in the denominator:

$$x^2 + 4x + 9 = (x + 2)^2 + 5.$$

* Integrals of types I and II can be easily integrated by means of the substitution $t = x - \alpha$.

This representation makes a substitution $x+2=t$, whence we get $x=t-2$, $dx=dt$, and therefore

$$\int \frac{6x+5}{x^2+4x+9}\,dx=\int \frac{6x+5}{(x+2)^2+5}\,dx=\int \frac{6(t-2)+5}{t^2+5}\,dt=\int \frac{6t-7}{t^2+5}\,dt$$

$$=3\int \frac{2t\,dt}{t^2+5}-7\int \frac{dt}{t^2+5}=3\ln(t^2+5)-\frac{7}{\sqrt{5}}\arctan\frac{t}{\sqrt{5}}+C.$$

Returning to the variable $x$, we obtain

$$\int \frac{6x+5}{x^2+4x+9}\,dx=3\ln(x^2+4x+9)-\frac{7}{\sqrt{5}}\arctan\frac{x+2}{\sqrt{5}}+C.$$

We shall now calculate the integral of type IV: $\int \frac{Ax+B}{(x^2+2px+q)^r}\,dx$, $q-p^2>0$, $r>1$. For that purpose, we introduce a new variable $z$ by the formula

$$z=(x+p)/\sqrt{q-p^2},$$

whence

$$x=z\sqrt{q-p^2}-p,\quad dx=\sqrt{q-p^2}\,dz. \tag{5}$$

Next we have

$$z^2+1=\frac{(x+p)^2}{q-p^2}+1=\frac{x^2+2px+q}{q-p^2}. \tag{6}$$

Thus, using substitution (5) and taking (6) into account, we obtain

$$\int \frac{Ax+B}{(x^2+2px+q)^r}\,dx=\int \frac{A\left(z\sqrt{q-p^2}-p\right)+B}{(z^2+1)^r(q-p^2)^r}\sqrt{q-p^2}\,dz$$
$$=\int \frac{Mz+N}{(z^2+1)^r}\,dz=M\int \frac{z\,dz}{(z^2+1)^r}+N\int \frac{dz}{(z^2+1)^r},$$

where $M$ and $N$ are numbers whose values are clear from the last but one relation.

We can apply the recursion formula (see formula (3) in 1.7.4, Item 3) to the second integral. Setting $z^2+1=t$ in the first integral, we obtain

$$M\int \frac{z\,dz}{(z^2+1)^r}=\frac{M}{2}\int \frac{dt}{t^r}=-\frac{M}{2(r-1)}\,\frac{1}{t^{r-1}}+C$$
$$=-\frac{M}{2(r-1)}\,\frac{1}{(z^2+1)^{r-1}}+C.$$

**Example 4.** Calculate $\int \frac{5x+3}{(x^2-2x+5)^2}\,dx$.

*Solution*. We set $z=\frac{x-1}{\sqrt{5-1}}=\frac{x-1}{2}$, whence $x=1+2z$, $dx=2\,dz$, $x^2-2x+5=4(z^2+1)$ and, consequently,

$$\int\frac{5x+3}{(x^2-2x+5)^2}dx=\int\frac{5(1+2z)+3}{4^2(z^2+1)^2}2dz=\int\frac{10z+8}{8(z^2+1)^2}dz$$
$$=\frac{5}{4}\int\frac{z\,dz}{(z^2+1)^2}+\int\frac{dz}{(z^2+1)^2}.$$

But

$$\int\frac{z\,dz}{(z^2+1)^2}=-\frac{1}{2}\,\frac{1}{z^2+1},\quad\int\frac{dz}{(z^2+1)^2}=\frac{z}{2(z^2+1)}+\frac{1}{2}\arctan z.$$

Thus

$$\int\frac{5x+3}{(x^2-2x+5)^2}dx=-\frac{5}{8}\,\frac{1}{z^2+1}+\frac{z}{2(z^2+1)}+\frac{1}{2}\arctan z+C$$
$$=\frac{4x-5}{8(z^2+1)}+\frac{1}{2}\arctan z+C.$$

Returning to the variable $x$, we obtain

$$\int\frac{5x+3}{(x^2-2x+5)^2}dx=\frac{2x-7}{2(x^2-2x+5)}+\frac{1}{2}\arctan\left(\frac{x-1}{2}\right)+C.$$

We have thus established that the integration of any rational function reduces to the integration of a polynomial and a finite number of partial fractions whose integrals can be expressed in terms of rational functions, logarithms and arc tangents. In other words, any rational function is integrable in elementary functions.

### 1.7.6. Integration of Irrational and Transcendental Functions

We shall first introduce the designation of a rational function of two variables $u$ and $v$, i.e. a function resulting from two variables $u$ and $v$ and some constants which are only involved in the operations of addition, subtraction, multiplication and division: $R(u, v)$. Such, for instance, is the function

$$R(u, v)=\frac{3u^2v+u^5v^4}{u^3+4v^3}.$$

If, in their turn, the variables $u$ and $v$ are functions of the variable $x$, i.e. $u=\varphi(x)$ and $v=\psi(x)$, then the function $R[\varphi(x), \psi(x)]$ is a rational function of $\varphi(x)$ and $\psi(x)$. For example, the function

$$f(x)=\frac{x+\sqrt{(x^2-1)^3}}{x^3-5\sqrt{x^2-1}}$$

is a rational function of $x$ and of $\sqrt{x^2-1}$: $f(x)=R(x, \sqrt{x^2-1})$. Here $R(u, v)=\frac{u+v^2}{u^2-5v}$, $u=x$, $v=\sqrt{x^2-1}$, and the function

$$f(x)=\frac{\sin^2 x-\cos^2 x}{\sin^2 x+2\cos x}$$

is a rational function of $\sin x$ and of $\cos x$: $f(x)=R(\sin x, \cos x)$.

We shall consider now the integrals of some simple irrational and transcendental functions and show that in some cases they reduce to the integrals of rational functions (or, as they say, are rationalized) and can be calculated by the methods discussed in 1.7.5.

**1. Integral of the form $\int R\left(x, \sqrt[m]{\frac{ax+b}{cx+d}}\right) dx$, where $a$, $b$, $c$, $d$ are some numbers $\left(\frac{a}{c}\neq\frac{b}{d}\right)$, $m$ is a natural number and $R$ is a rational function of $x$ and of $\sqrt[m]{\frac{ax+b}{cx+d}}$. We shall show that such an integral can be rationalized by means of a substitution**

$$t=\sqrt[m]{\frac{ax+b}{cx+d}}.$$

Indeed,

$$t^m=\frac{ax+b}{cx+d},\ x=\frac{b-dt^m}{ct^m-a},\ dx=\frac{mt^{m-1}(ad-bc)}{(ct^m-a)^2}\,dt$$

so that

$$\int R\left(x, \sqrt[m]{\frac{ax+b}{cx+d}}\right) dx=\int R\left(\frac{b-dt^m}{ct^m-a}, t\right)\frac{mt^{m-1}(ad-bc)}{(ct^m-a)^2}\,dt=\int R_1(t)\,dt,$$

where $R_1(t)$ is a rational function of the argument $t$.

**Example 1.** Calculate $\int \sqrt{\frac{1+x}{1-x}}\,\frac{dx}{1-x}$.

*Solution.* Making the substitution $t=\sqrt{\frac{1+x}{1-x}}$, we get $t^2=\frac{1+x}{1-x}$, $1-x=\frac{2}{t^2+1}$, $x=\frac{t^2-1}{t^2+1}$, $dx=\frac{4t\,dt}{(t^2+1)^2}$. Next we have

$$\int \sqrt{\frac{1+x}{1-x}}\,\frac{dx}{1-x}=2\int\frac{t^2\,dt}{t^2+1}=2\int\frac{t^2+1-1}{t^2+1}\,dt=2\int dt-2\int\frac{dt}{t^2+1}$$

$$=2t-2\arctan t+C=2\sqrt{\frac{1+x}{1-x}}-2\arctan\sqrt{\frac{1+x}{1-x}}+C.$$

**Example 2.** Calculate $\int\frac{dx}{\sqrt{x}+\sqrt[3]{x}}$.

*Solution.* We have

$$\int \frac{dx}{\sqrt{x}+\sqrt[3]{x}} = \int \frac{dx}{(\sqrt[6]{x})^2+(\sqrt[6]{x})^3} = \left| \begin{matrix} t=\sqrt[6]{x},\ x=t^6 \\ dx=6t^5\,dt \end{matrix} \right| = \int \frac{6t^5\,dt}{t^2+t^3}$$

$$=6\int \frac{t^3\,dt}{t+1} = 6\int \left[(t^2-t+1)\,dt - \int \frac{dt}{t+1}\right]$$

$$=6\left(\frac{t^3}{3}-\frac{t^2}{2}+t-\ln|t+1|\right)+C$$

$$=2\sqrt{x}-3\sqrt[3]{x}+6\sqrt[6]{x}-6\ln(\sqrt[6]{x}+1)+C.$$

**2. Integral of the form $\int R(x, \sqrt{ax^2+bx+c})\,dx$, where $a$, $b$, $c$ are some numbers, $a \neq 0$, and $R$ is a rational function of $x$ and of $\sqrt{ax^2+bx+c}$.**

If the trinomial $ax^2+bx+c$ has real roots $x_1$, $x_2$ ($x_1 \neq x_2$) and $a>0$, then

$$\sqrt{ax^2+bx+c} = \sqrt{a(x-x_1)(x-x_2)} = |x-x_1|\sqrt{\frac{a(x-x_2)}{x-x_1}}.$$

Consequently,

$$R(x, \sqrt{ax^2+bx+c}) = R\left(x, |x-x_1|\sqrt{\frac{a(x-x_2)}{x-x_1}}\right)$$
$$= R_1\left(x, \sqrt{\frac{a(x-x_2)}{x-x_1}}\right),$$

i.e. we get an integral considered in Item 1.

If $x_1 = x_2$, then

$$\sqrt{ax^2+bx+c} = |x-x_1|\sqrt{a},$$

i.e. a rational function of $x$ is under the integral sign.

Therefore of interest is the case when the trinomial $ax^2+bx+c$ does not have any real roots and $a>0$. We shall show that in this case the integral can be rationalized by means of Euler's* substitution

$$t=\sqrt{ax^2+bx+c}+x\sqrt{a}.^{**}$$

---

* Euler, Leonhard (1707-1783), an eminent mathematician, mechanician, physicist and astronomer. Member of the Petersburg Academy of Sciences; a Swiss by nationality, most of his life lived in Russia.

** We can set $t=\sqrt{ax^2+bx+c}-x\sqrt{a}$.

Squaring both sides of the relation $\sqrt{ax^2+bx+c}=t-x\sqrt{a}$, we get $bx+c=t^2-2\sqrt{a}\,tx$ so that

$$x=\frac{t^2-c}{2\sqrt{a}\,t+b},\quad \sqrt{ax^2+bx+c}=\frac{\sqrt{a}\,t^2+bt+c\sqrt{a}}{2\sqrt{a}\,t+b},$$

$$\mathrm{d}x=2\,\frac{\sqrt{a}\,t^2+bt+c\sqrt{a}}{(2\sqrt{a}\,t+b)^2}\,\mathrm{d}t.$$

Thus

$$\int R(x,\sqrt{ax^2+bx+c})\,\mathrm{d}x$$

$$=\int R\left(\frac{t^2-c}{2\sqrt{a}\,t+b},\ \frac{\sqrt{a}\,t^2+bt+c\sqrt{a}}{2\sqrt{a}\,t+b}\right)2\,\frac{\sqrt{a}\,t^2+bt+c\sqrt{a}}{(2\sqrt{a}\,t+b)^2}\,\mathrm{d}t=\int R_1(t)\,\mathrm{d}t,$$

where $R_1(t)$ is a rational function of $t$.

Now if $a<0$ in the trinomial $ax^2+bx+c$ and $c>0$, then we can use another Euler's substitution to rationalize the integral:

$$\sqrt{ax^2+bx+c}=xt\pm\sqrt{c}.$$

**Example 3.** Calculate $\int\frac{\mathrm{d}x}{x+\sqrt{x^2+x+1}}$.

*Solution.* Since the trinomial $x^2+x+1$ has complex roots, we make a substitution $\sqrt{x^2+x+1}=t-x$. Squaring both sides of the relation, we obtain $x^2+x+1=t^2-2tx+x^2$, or $x+1=t^2-2tx$, whence we get

$$x=\frac{t^2-1}{1+2t},\quad \mathrm{d}x=2\,\frac{t^2+t+1}{(1+2t)^2}\,\mathrm{d}t.$$

Then

$$\int\frac{\mathrm{d}x}{x+\sqrt{x^2+x+1}}=2\int\frac{t^2+t+1}{t(1+2t)^2}\,\mathrm{d}t.$$

Furthermore we have

$$\frac{2t^2+2t+2}{t(1+2t)^2}=\frac{A}{t}+\frac{B}{1+2t}+\frac{D}{(1+2t)^2}.$$

Multiplying both sides of the relation by $t(1+2t)^2$, we obtain

$$2t^2+2t+2=A(1+2t)^2+Bt(1+2t)+Dt,$$

or

$$2t^2+2t+2=(4A+2B)t^2+(4A+D+B)t+A.$$

Equating the coefficients in the equal powers of $t$ to one another, we get a system of first-degree equations with respect to $A$, $B$, $D$:

$$\begin{cases} 4A+2B=2, \\ 4A+D+B=2, \\ A=2, \end{cases}$$

whence $A=2$, $B=-3$, $D=-3$. Consequently,

$$\frac{2t^2+2t+2}{t(1+2t)^2}=\frac{2}{t}-\frac{3}{1+2t}-\frac{3}{(1+2t)^2},$$

and the final result is

$$\int\frac{dx}{x+\sqrt{x^2+x+1}}=\int\left[\frac{2}{t}-\frac{3}{1+2t}-\frac{3}{(1+2t)^2}\right]dt$$

$$=2\int\frac{dt}{t}-\frac{3}{2}\int\frac{d(1+2t)}{1+2t}-\frac{3}{2}\int\frac{d(1+2t)}{(1+2t)^2}$$

$$=2\ln|t|-\frac{3}{2}\ln|1+2t|+\frac{3}{2(1+2t)}+C=2\ln\left|\sqrt{x^2+x+1}+x\right|$$

$$-\frac{3}{2}\ln\left|1+2x+2\sqrt{x^2+x+1}\right|+\frac{3}{2(1+2x+2\sqrt{x^2+x+1})}+C.$$

**Example 4.** Calculate $\int\frac{dx}{(1+x)\sqrt{1+x-x^2}}$.

*Solution.* Here the trinomial $1+x-x^2$ has complex roots, $a<0$, $c>0$, and therefore we use the substitution $\sqrt{1+x-x^2}=tx-1$. Squaring both sides of the relation, we get $1+x-x^2=t^2x^2-2tx+1$, or $1-x=t^2x-2t$, whence we get

$$x=\frac{1+2t}{t^2+1},\quad dx=\frac{2(1-t-t^2)}{(t^2+1)^2}dt,\quad \sqrt{1+x-x^2}=\frac{t^2+t-1}{t^2+1}.$$

Thus

$$\int\frac{dx}{(1+x)\sqrt{1+x-x^2}}=\int\frac{2(1-t-t^2)}{\left(1+\frac{1+2t}{t^2+1}\right)\frac{t^2+t-1}{t^2+1}\,(t^2+1)^2}dt$$

$$=-2\int\frac{dt}{1+(t+1)^2}=-2\arctan(t+1)+C$$

$$=-2\arctan\frac{\sqrt{1+x-x^2}+x+1}{x}+C.$$

Note that calculations of integrals with the use of Euler's substitutions usually lead to cumbersome expressions and laborious computations. Therefore they should only be used if it is impossible to calculate the integral by a simpler technique.

**3. Integral of the form $\int R(\sin x, \cos x)\,dx$, where $R$ is a rational function of $\sin x$ and of $\cos x$.** We shall show that the integral can

be rationalized by means of the substitution

$$t = \tan \frac{x}{2}, \quad -\pi < x < \pi.$$

Indeed,

$$\sin x = \frac{2\tan(x/2)}{1+\tan^2(x/2)} = \frac{2t}{1+t^2}, \quad \cos x = \frac{1-\tan^2(x/2)}{1+\tan^2(x/2)} = \frac{1-t^2}{1+t^2},$$

$$x = 2\arctan t, \quad dx = \frac{2dt}{1+t^2}$$

so that

$$\int R(\sin x, \cos x)\,dx = \int R\left(\frac{2t}{1+t^2}, \frac{1-t^2}{1+t^2}\right)\frac{2dt}{1+t^2} = \int R_1(t)\,dt,$$

where $R_1(t)$ is a rational function of $t$.

**Example 5.** Calculate $\int \frac{dx}{1+\sin x}$.

*Solution.* Using the substitution $t = \tan(x/2)$, we obtain

$$\sin x = \frac{2t}{1+t^2}, \quad x = 2\arctan t, \quad dx = \frac{2dt}{1+t^2}.$$

Thus

$$\int \frac{dx}{1+\sin x} = 2\int \frac{dt}{(1+t)^2} = -\frac{2}{1+t} + C = -\frac{2}{1+\tan(x/2)} + C.$$

**4. Integral of the form $\int R(e^x)\,dx$.** We shall show that this integral can be rationalized by means of the substitution

$$t = e^x.$$

Indeed, since $x = \ln t$ and $dx = dt/t$, it follows that

$$\int R(e^x)\,dx = \int R(t)\frac{dt}{t},$$

where $R(t)$ is a rational function of $t$.

**Example 6.** Calculate $\int \frac{e^x-1}{e^x+1}\,dx$.

*Solution.* We set $t = e^x$, whence it follows that $dx = dt/t$. Consequently,

$$\int \frac{e^x-1}{e^x+1}\,dx = \int \frac{t-1}{t+1}\frac{dt}{t} = \int \frac{2t-(t+1)}{(t+1)t}\,dt = 2\int \frac{dt}{t+1} - \int \frac{dt}{t}$$
$$= 2\ln(1+t) - \ln t + C = 2\ln(1+e^x) - x + C.$$

It should be pointed out in conclusion that the methods and techniques of integration we have considered do not exhaust all classes of analytically integrable elementary functions. At the same time, it follows from what we have discussed that integration is more complicated than differentiation. Definite skills and inventiveness

are needed which are acquired in practice as a result of solving a large number of problems.

It should also be noted that whereas differentiation does not take us outside the class of elementary functions, it is not so in integration. There are elementary functions (say, $e^{-x^2}$, $1/(\ln x)$, $(\sin x)/x$ and others) whose antiderivatives are not elementary functions. Such antiderivatives not only exist but also play a significant part in analysis itself and in its applications. They are well known, tables and graphs are constructed for them, which help using them in practical application.

If an antiderivative is not an elementary function, then we say that the integral "cannot be taken" in elementary functions.

## Chapter 8

## THE DEFINITE INTEGRAL

### 1.8.1. Definition of the Definite Integral

Let the function $y = f(x)$ be defined on the closed interval $[a, b]$, $a < b$. We divide the interval into $n$ arbitrary parts by points:

$$a = x_0 < x_1 < x_2 < \ldots < x_{i-1} < x_i < \ldots < x_n = b.$$

We designate this division as $\tau$ and the points $x_0, x_1, \ldots, x_n$ as the points of division. In each of the resulting subintervals $[x_{i-1}, x_i]$ we choose an arbitrary point $\xi_i$ $(x_{i-1} \leqslant \xi_i \leqslant x_i)$. We designate as $\Delta x_i$ the difference $x_i - x_{i-1}$ which we agree to call the length of the subinterval $[x_{i-1}, x_i]$. We form a sum

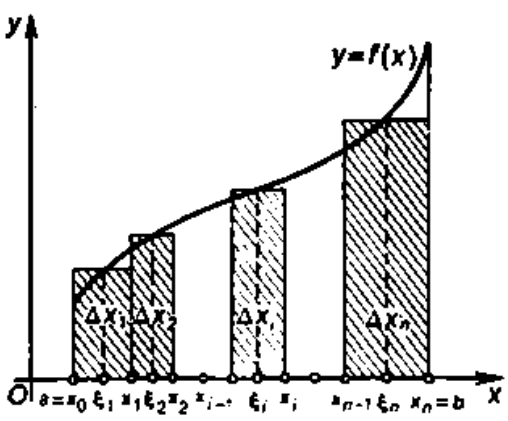



Fig. 93

$$\sigma = f(\xi_1)\,\Delta x_1 + f(\xi_2)\,\Delta x_2 + \ldots + f(\xi_n)\Delta x_n = \sum_{i=1}^{n} f(\xi_i)\,\Delta x_i, \qquad (1)$$

which we shall call the *integral sum* for the function $f(x)$ on $[a,b]$ corresponding to the given division of $[a, b]$ into subintervals and to the given choice of the intermediate points $\xi_i$. The geometric meaning of the sum $\sigma$ is obvious: it is the sum of the areas of rectangles with the bases $\Delta x_1, \Delta x_2, \ldots, \Delta x_n$ and altitudes $f(\xi_1), f(\xi_2), \ldots, f(\xi_n)$ (Fig. 93) (provided that $f(x) \geqslant 0$).

We designate the length of the largest subinterval $\tau$ as $\lambda$: $\lambda = \max\limits_{1 \leqslant i \leqslant n} \{\Delta x_i\}$.

**Definition.** *If there is a finite limit $I$ of the integral sum* (1) *as $\lambda \to 0$, then this limit is the definite integral of the function $f(x)$ with respect to the closed interval $[a, b]$ and is designated as follows:*

$$I = \int_a^b f(x)\,dx, \tag{2}$$

*or*

$$\int_a^b f(x)\,dx = \lim_{\lambda \to 0} \sum_{i=1}^{n} f(\xi_i)\,\Delta x_i.$$

In this case the function $f(x)$ is said to be *integrable* on $[a, b]$. The numbers $a$ and $b$ are the *lower* and the *upper limit of integration* respectively, $f(x)$ is the *integrand* and $x$ is the *integration variable*.

Some explanations are due since here is an unusual limiting process. Indeed, the integral sum depends on the points of division $x_i$ and the intermediate points $\xi_i$. The number of both kinds of points tends to infinity as $\lambda \to 0$. Therefore the concept of the limit of an integral sum must be explained.

We shall first give the corresponding definition "in the language of sequences". Assume that we divide successively the interval $[a, b]$ into parts first by one technique, then by another technique, then by some other technique, and so on, the length $\lambda_k$ of the largest subinterval which results from the $k$th division tending to zero as $k$ tends to infinity. In each division we choose arbitrarily the intermediate points $\xi_i$. We thus obtain a sequence of divisions $\{\tau_k\}$ in which $\lim\limits_{k \to \infty} \lambda_k = 0$ and can give a definition of the definite integral "in the language of sequences": *the function $f(x)$ is integrable on the closed interval $[a, b]$ if for any sequence of divisions $\{\tau_k\}$ in which $\lim\limits_{k \to \infty} \lambda_k = 0$ the corresponding sequence of integral sums $\{\sigma_k\}$ tends to the same number $I$.*

We can also give a definition of the definite integral "in the $\varepsilon$-$\delta$ language": *the number $I$ is the definite integral of the function $f(x)$ with respect to the closed interval $[a, b]$ if for any $\varepsilon > 0$ there is $\delta > 0$ such that for $\lambda < \delta$ (i.e. if the interval is divided into parts $\Delta x_i < \delta$ long), irrespective of the choice of the points $\xi_i$, the inequality*

$$\left| \sum_{i=1}^{n} f(\xi_i)\,\Delta x_i - I \right| < \varepsilon$$

*is satisfied.*

The proof of the equivalence of the two definitions is similar to the proof of the equivalence of two definitions of the limit of a func-

tion. The definition "in the language of sequences" makes it possible to extend the main concepts of the theory of limits to this new kind of a limit.

From the definition of the definite integral it follows that the value of integral (2) depends only on the form of the function $f(x)$ and on the numbers $a$ and $b$. Consequently, if $f(x)$ and the limits of integration are defined, then integral (2) is uniquely defined and is a number. It follows, in particular, that the definite integral does not depend on the choice of the designation of the argument of the integrand, i.e. on the designation of the integration variable:

$$\int_a^b f(x)\,dx = \int_a^b f(t)\,dt = \int_a^b f(\xi)\,d\xi \text{ etc.}$$

### 1.8.2. The Conditions for the Existence of the Definite Integral

**1. Boundedness of the integrable function.**

**Theorem 8.1 (the necessary condition for the integrability of a function).** *If the function $f(x)$ is integrable on the closed interval $[a, b]$, then it is bounded on this interval.*

*Proof.* We assume the contrary, i.e. let the function $f(x)$ be unbounded on $[a, b]$. We shall show that in this case, by properly choosing points $\xi_1, \xi_2, \ldots, \xi_n$, we can make the integral sum $\sigma$ arbitrarily large for any division of the interval $[a, b]$.

Indeed, since $f(x)$ is unbounded on $[a, b]$, for any division of the interval $[a, b]$ it possesses this property at least on one of the subintervals, say, on $[x_0, x_1]$. We choose arbitrarily points $\xi_2, \xi_3, \ldots, \xi_n$ on the other subintervals and introduce the designation

$$\sigma' = f(\xi_2)\,\Delta x_2 + f(\xi_3)\,\Delta x_3 + \ldots + f(\xi_n)\,\Delta x_n.$$

We take an arbitrary number $M > 0$ and choose $\xi_1$ on $[x_0, x_1]$ such that

$$|f(\xi_1)| \geqslant \frac{|\sigma'| + M}{\Delta x_1}.$$

We can do this by virtue of the unboundedness of the function $f(x)$ on $[x_0, x_1]$. Then

$$|f(\xi_1)|\,\Delta x_1 \geqslant |\sigma'| + M,$$

$$|\sigma| = |f(\xi_1)\Delta x_1 + \sigma'| \geqslant |f(\xi_1)|\,\Delta x_1 - |\sigma'| \geqslant M,$$

i.e. the integral sum $\sigma$ is larger in absolute value than any of the preassigned numbers. Therefore the integral sum $\sigma$ does not possess a finite limit as $\lambda \to 0$, and this means that there is no definite integral of an unbounded function. ■

**Remark.** The converse theorem does not hold true, i.e. the condition for the boundedness of the function $f(x)$ is a necessary but not a sufficient condition for the integrability of the function. Here is an example to elucidate this statement. We consider Dirichlet's function on the interval [0, 1]:

$$f(x)=\begin{cases}1 & \text{if } x \text{ is a rational number,}\\ 0 & \text{if } x \text{ is an irrational number.}\end{cases}$$

Dirichlet's function is evidently bounded. However, it is not integrable on [0, 1]. We shall show this. If, for any division of the interval [0, 1], we choose rational points $\xi_i$ $(x_{i-1} \leqslant \xi_i \leqslant x_i)$, then we obtain

$$\sigma=\sum_{i=1}^{n} f(\xi_i)\,\Delta x_i=\sum_{i=1}^{n} 1\cdot\Delta x_i=1,$$

and if we take irrational points $\xi_i$, we get

$$\sigma=\sum_{i=1}^{n} f(\xi_i)\,\Delta x_i=\sum_{i=1}^{n} 0\cdot\Delta x_i=0.$$

Thus, upon a division into arbitrarily small subintervals, the integral sum can assume a value equal to 0 and a value equal to 1. Therefore, as $\lambda \to 0$, the integral sum $\sigma$ has no limit.

Thus, for a definite integral of a function $f(x)$ to exist, the function $f(x)$ must be bounded and possess other properties ensuring its integrability. To find these properties, we must introduce the concepts of the lower and upper sums.

**2. Darboux'* sums.** Assume that the function $f(x)$ is bounded on the closed interval $[a, b]$ and $\tau$ is the division of the interval by points: $a = x_0 < x_1 < \ldots < x_{i-1} < x_i < \ldots < x_n = b$. We designate as $m_i$ and $M_i$ the greatest lower bound and the least upper bound of this function on the subinterval $[x_{i-1}, x_i]$ respectively and form the following sums:

$$S=M_1\Delta x_1+M_2\Delta x_2+\ldots+M_n\Delta x_n=\sum_{i=1}^{n} M_i\,\Delta x_i,$$

$$s=m_1\Delta x_1+m_2\Delta x_2+\ldots+m_n\Delta x_n=\sum_{i=1}^{n} m_i\,\Delta x_i.$$

These sums are called the *upper* and the *lower sum*, or the *upper Darboux sum* and the *lower Darboux sum*, respectively, of the function $f(x)$ for the given division $\tau$ of the interval $[a, b]$.

* Darboux, Jean Gaston (1842-1917), a French mathematician.

From the definition of the lower and the upper bound it follows that $m_i \leqslant f(\xi_i) \leqslant M_i$ for $\xi_i \in [x_{i-1}, x_i]$. Hence

$$s = \sum_{i=1}^{n} m_i \Delta x_i \leqslant \sigma = \sum_{i=1}^{n} f(\xi_i) \Delta x_i \leqslant \sum_{i=1}^{n} M_i \Delta x_i = S,$$

i.e. any integral sum and Darboux' sums for the given division are related as

$$s \leqslant \sigma \leqslant S. \tag{1}$$

Darboux' sums have a simple geometric meaning. Let us consider a nonnegative continuous function $f(x)$ on $[a, b]$ and a curvilinear trapezoid, i.e. a figure bounded by the graph of the function $f(x)$, by two vertical straight lines drawn through the points $a$ and $b$

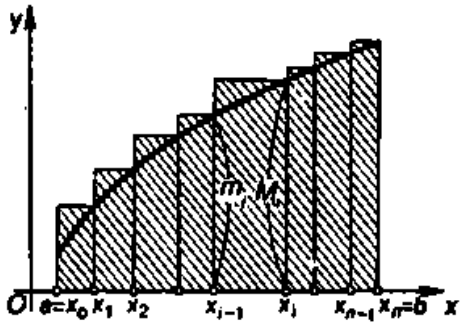



Fig. 94

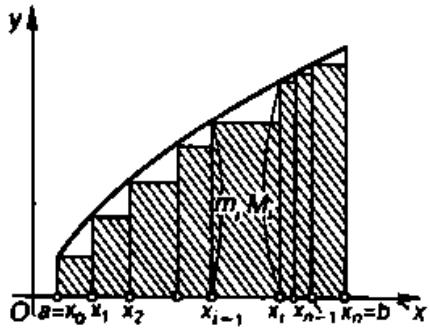



Fig. 95

of the $x$-axis and by the $x$-axis (Figs. 94 and 95). Since the function $f(x)$ is continuous on $[a, b]$, it is also continuous on $[x_{i-1}, x_i]$. According to Weierstrass's second theorem, the function $f(x)$ attains its greatest lower bound and least upper bound on $[x_{i-1}, x_i]$ and, consequently, $m_i$ and $M_i$ are the least and the greatest value of the function on this interval respectively. Therefore the sum $S$ is equal to the area of the stepped figure hatched in Fig. 94 which is "circumscribed" about the curvilinear trapezoid, and the sum $s$ is equal to the area of the stepped figure hatched in Fig. 95 which is "inscribed" into the given curvilinear trapezoid.

It should be specially pointed out that Darboux' sums depend only on the division of the closed interval $[a, b]$, whereas the integral sum $\sigma$ also depends on the choice of points $\xi_i$ on the subintervals $[x_{i-1}, x_i]$. For a fixed division of the interval $[a, b]$ the sums $s$ and $S$ are arbitrary numbers and the sum $\sigma$ is a variable quantity since the points $\xi_i$ are arbitrary.

**3. Properties of Darboux' sums.**

1°. *For any fixed division $\tau$ and for any $\varepsilon > 0$ we can choose points $\xi_i$ on the subintervals $[x_{i-1}, x_i]$ such that the integral sum $\sigma$ satisfies the inequalities $0 \leqslant S - \sigma < \varepsilon$. The points $\xi_i$ can also be chosen such that the integral sum satisfies the inequalities $0 \leqslant \sigma - s < \varepsilon$.*

*Proof.* Let $\tau$ be a fixed division of the closed interval $[a, b]$. We shall prove, for instance, the inequalities $0 \leqslant S - \sigma < \varepsilon$. According to the property of the least upper bound $M_i$, for the given $\varepsilon > 0$, on the subinterval $[x_{i-1}, x_i]$ we can indicate a point $\xi_i$ such that

$$0 \leqslant M_i - f(\xi_i) < \frac{\varepsilon}{b-a}, \quad i = 1, 2, \ldots, n.$$

Multiplying these inequalities by $\Delta x_i$ and then adding the results together, we get $0 \leqslant S - \sigma < \varepsilon$.

We can similarly establish the inequalities $0 \leqslant \sigma - s < \varepsilon$. ■

2°. *When new division points are added to the given division $\tau$ of the closed interval $[a, b]$, the lower Darboux sum does not decrease and the upper Darboux sum does not increase.*

*Proof.* It is sufficient to add one more division point $x'$ to the division $\tau$ since we can add several division points one by one. Assume that the new division point $x'$ falls on the subinterval $[x_{i-1}, x_i]$

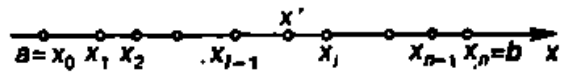



Fig. 96

(Fig. 96). We designate the lower Darboux sums as $s$ and $s'$ and the upper Darboux sums as $S$ and $S'$ for the division $\tau$ and for the division $\tau'$ obtained from it by adding the point $x'$.

We shall carry out the proof for the lower Darboux sums $s$ and $s'$. We designate as $m_i'$ and $m_i''$ the greatest lower bounds of the function $f(x)$ on the subintervals $[x_{i-1}, x']$ and $[x', x_i]$ respectively. The sum $s$ includes the summand $m_i \Delta x_i$, and the sum $s'$ includes the summands $m_i'(x' - x_{i-1}) + m_i''(x_i - x')$ instead of it. The other summands are the same in the sums $s$ and $s'$. Since $m_i' \geqslant m_i$, $m_i'' \geqslant m_i$ (the greatest lower bound on the subinterval $[x_{i-1}, x_i]$ is not smaller than the greatest lower bound on the whole subinterval $[x_{i-1}, x_i]$), it follows that

$$m_i'(x' - x_{i-1}) + m_i''(x_i - x') \geqslant m_i(x' - x_{i-1}) + m_i(x_i - x') = m_i \Delta x_i.$$

Hence $s' \geqslant s$.

We can prove by analogy that $S' \leqslant S$. ■

3°. *The lower Darboux sum for any division $\tau'$ does not exceed the upper Darboux sum for any other division $\tau''$.*

*Proof.* Let $s'$ and $S'$, $s''$ and $S''$ be the lower and the upper Darboux sum for the divisions $\tau'$ and $\tau''$ respectively. We consider the division $\tau$ consisting of all points which belong to the divisions $\tau'$ and $\tau''$, and designate its Darboux sums as $s$ and $S$. Since the divi-

sion $\tau$ can be obtained from the division $\tau'$ by adding to it the division points of $\tau''$, it follows, according to Property 2°, with due account of the obvious inequality $s \leqslant S$, that

$$s' \leqslant s \leqslant S \leqslant S'.$$

But we can also obtain the division $\tau$ from the division $\tau''$ by adding the division points of $\tau'$. Therefore

$$s'' \leqslant s \leqslant S \leqslant S''.$$

Comparing the inequalities we have established, we find that $s' \leqslant S''$ and $s'' \leqslant S'$.■

4°. *The set $\{S\}$ of the upper Darboux sums of the function $f(x)$ for various divisions of the closed interval $[a, b]$ is bounded from below and the set $\{s\}$ of the lower Darboux sums is bounded from above, and the least upper bound of the set $\{s\}$ does not exceed the greatest lower bound of the set $\{S\}$.*

*Proof.* This property follows directly from Property 3°. Indeed, the set $\{S\}$ of all upper Darboux sums is bounded from below, say, by any lower Darboux sum $s$, and the set $\{s\}$ of all lower Darboux sums is bounded from above, say, by any upper Darboux sum $S$. Therefore, in accordance with Theorem 1.1, the sets $\{S\}$ and $\{s\}$ have the least upper bound and the greatest lower bound. We designate the greatest lower bound of the set $\{S\}$ as $I^*$ and the least upper bound of the set $\{s\}$ as $I_*$:

$$I^* = \inf \{S\}, \ I_* = \sup \{s\}.$$

We shall show that $I_* \leqslant I^*$. Let $I_* > I^*$. We designate their difference as $\varepsilon$ so that $I_* - I^* = \varepsilon > 0$. From the property of the supremum $I_*$ and infimum $I^*$ it follows that there are numbers $S'$ and $s''$ which are the upper and the lower Darboux sum respectively of divisions $\tau'$ and $\tau''$ of the closed interval $[a, b]$ such that $I^* + \varepsilon/2 > S'$ and $I_* - \varepsilon/2 < s''$. Subtracting the second inequality from the first, we get $S' - s'' < I^* - I_* + \varepsilon$. But $I^* - I_* = -\varepsilon$ and therefore $S' - s'' < 0$, i.e. $s'' > S'$, and this contradicts Property 3°. Consequently, $I_* \leqslant I^*$.■

**4. The necessary and sufficient condition for integrability.**

Let us consider the following fundamental theorem.

**Theorem 8.2.** *For the function $f(x)$ bounded on the closed interval $[a, b]$ to be integrable on this interval, it is necessary and sufficient that*

$$\lim_{\lambda \to 0} (S - s) = 0. \tag{2}$$

*Condition* (2) *means that for any $\varepsilon > 0$ there is $\delta > 0$ such that the inequality $|S - s| < \varepsilon$ holds true for $\lambda < \delta$. Since $s \leqslant S$, the last*

*inequality is equivalent to the inequality*

$$S - s < \varepsilon. \tag{3}$$

*Proof*. **Necessity.** Let the function $f(x)$ be integrable on the closed interval $[a, b]$, i.e. there is a definite integral $I = \int_a^b f(x)\,dx$. This means that for any $\varepsilon > 0$ there is $\delta > 0$ such that the inequality

$$|\sigma - I| < \frac{\varepsilon}{4} \tag{4}$$

holds true for any division $\tau$ which satisfies the condition $\lambda < \delta$ irrespective of the choice of the points $\xi_i$. We fix any division $\tau$ of this kind. According to Property 1°, there are integral sums $\sigma'$ and $\sigma''$ for it such that

$$S - \sigma' \leqslant \frac{\varepsilon}{4}, \quad \sigma'' - s \leqslant \frac{\varepsilon}{4}. \tag{5}$$

Note that both integral sums $\sigma'$ and $\sigma''$ satisfy inequality (4). From the relation

$$S - s = (S - \sigma') + (\sigma' - I) + (I - \sigma'') + (\sigma'' - s)$$

and inequalities (4) and (5) it follows that

$$S - s < \varepsilon,$$

and this means that condition (3) is fulfilled.

**Sufficiency.** Assume that condition (3) is fulfilled. According to Property 4°, $s \leqslant I_* \leqslant I^* \leqslant S$ for any lower and upper Darboux sums, and therefore $0 \leqslant I^* - I_* \leqslant S - s$, whence, according to (3), it follows that $0 \leqslant I^* - I_* < \varepsilon$ for any $\varepsilon > 0$. This means that $I^* - I_* = 0$, i.e. $I^* = I_*$. Setting $I = I^* = I_*$, we find that the inequalities

$$s \leqslant I \leqslant S \tag{6}$$

hold true for any division. Now if the integral sum $\sigma$ and Darboux sums $s$ and $S$ correspond to the same division $\tau$, then, as is known (see inequality (1)),

$$s \leqslant \sigma \leqslant S. \tag{7}$$

It follows from inequalities (6) and (7) that

$$|\sigma - I| \leqslant S - s. \tag{8}$$

By the hypothesis, for any $\varepsilon > 0$ there is $\delta > 0$ such that for $\lambda < \delta$ inequality (3) holds true: $S - s < \varepsilon$. But then it follows from inequality (8) that

$$|\sigma - I| < \varepsilon \text{ for } \lambda < \sigma$$

as well, and this means that the number $I$ is a limit of the integral sum $\sigma$ as $\lambda \to 0$, i.e. the function $f(x)$ is integrable on the interval $[a, b]$. ■

In what follows, we shall need another notation for the necessary and sufficient condition for integrability. Designating the oscillation $M_i - m_i$ of the function $f(x)$ on the subinterval $[x_{i-1}, x_i]$ as $\omega_i$, we have

$$S - s = \sum_{i=1}^{n} M_i \Delta x_i - \sum_{i=1}^{n} m_i \Delta x_i = \sum_{i=1}^{n} (M_i - m_i) \Delta x_i = \sum_{i=1}^{n} \omega_i \Delta x_i.$$

Since $M_i \geqslant m_i$ and $\Delta x_i > 0$, each term in the last sum is nonnegative, and we can write the condition for the existence of a definite integral as follows: for any $\varepsilon > 0$ there is $\delta > 0$ such that

$$\sum_{i=1}^{n} \omega_i \Delta x_i < \varepsilon \quad \text{for} \quad \lambda < \delta.$$

It is the form in which it is usually used.

### 1.8.3. Integrability of Continuous and Certain Discontinuous Functions

**Theorem 8.3.** *If the function $f(x)$ is continuous on the closed interval $[a, b]$, then it is integrable on it.*

*Proof.* Since the function $f(x)$ is continuous on the closed interval $[a, b]$, it follows, according to Cantor's theorem, that it is uniformly continuous on it. Let $\varepsilon > 0$. According to the corollary of Cantor's theorem, for the positive number $\varepsilon/(b - a)$ there is $\delta > 0$ such that when the interval $[a, b]$ is divided into subintervals $[x_{i-1}, x_i]$, whose length is $\Delta x_i < \delta$, all oscillations $\omega_i$ are smaller than $\varepsilon/(b - a)$. Hence

$$S - s = \sum_{i=1}^{n} \omega_i \Delta x_i < \frac{\varepsilon}{b-a} \sum_{i=1}^{n} \Delta x_i = \varepsilon \text{ for } \lambda < \delta.$$

Consequently, for the function $f(x)$ continuous on the interval $[a, b]$ the sufficient condition for integrability is fulfilled and this condition implies the existence of a definite integral. ■

It follows from the theorem that the condition for the continuity of a function is the sufficient condition for its integrability. But this does not mean that a definite integral exists only for continuous functions. The class of integrable functions is much larger. Thus, for instance, there is a definite integral of functions which have a finite number of points of discontinuity. We shall prove this.

**Theorem 8.4.** *If the function $f(x)$ is bounded on the closed interval $[a, b]$ and continuous on it everywhere, except for a finite number of points, then it is integrable on this interval.*

*Proof.* It is sufficient to consider the case when there is only one point of discontinuity, $x'$, between $a$ and $b$. Let $M$ and $m$ be the supremum and infimum of the function $f(x)$ on $[a, b]$ and $\Omega = M - m$ be its oscillation on this interval. We take any sufficiently small $\varepsilon > 0$ and consider intervals $[a, x' - \varepsilon/(8\Omega)]$ and $[x' + \varepsilon/(8\Omega), b]$ (Fig. 97). On each of these intervals the function $f(x)$ is continuous

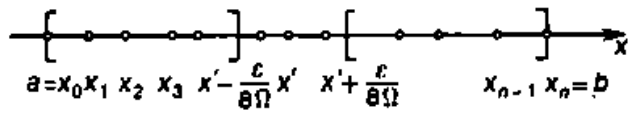



Fig. 97

and, consequently, there is $\delta' > 0$ such that when they are divided into subintervals $[x'_{i-1}, x'_i]$ of length $\Delta x'_i < \delta'$, all oscillations $\omega'_i$ are smaller than $\frac{\varepsilon}{2(b-a)}$.

Let $\delta = \min\{\delta', \varepsilon/(8\Omega)\}$. We shall consider now an arbitrary division of $[a, b]$ into subintervals whose length is $\Delta x_i < \delta$ (see Fig. 97). For this division we divide the sum $\sum_{i=1}^{n} \omega_i \Delta x_i$ into summands $\sum' \omega_i \Delta x_i + \sum'' \omega_i \Delta x_i$, where the first sum includes the subintervals which lie entirely outside the $\varepsilon/(8\Omega)$-neighbourhood of the point $x'$ and the second sum includes the subintervals which lie entirely inside the $\varepsilon/(8\Omega)$-neighbourhood of the point $x'$ or have points in common with it.

As in the proof of the preceding theorem, we have for the first sum

$$\sum{}' \omega_i \Delta x_i < \frac{\varepsilon}{2},$$

and, as concerns the second sum, we note that the sum of the lengths of the subintervals which are entirely inside the $\varepsilon/(8\Omega)$-neighbourhood of the point $x'$ is smaller than or equal to $\varepsilon/(4\Omega)$; the number of subintervals which only partly lie in this neighbourhood is not larger than two and therefore the sum of their lengths is smaller than $2\delta \leqslant \varepsilon/(4\Omega)$. Consequently,

$$\sum{}'' \omega_i \Delta x_i \leqslant \Omega \sum{}'' \Delta x_i < \Omega \frac{\varepsilon}{2\Omega} = \frac{\varepsilon}{2}.$$

Thus the final result is

$$S - s = \sum_{i=1}^{n} \omega_i \Delta x_i = \sum{}' \omega_i \Delta x_i + \sum{}'' \omega_i \Delta x_i < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \quad \text{for} \quad \lambda < \delta.$$

And this proves the integrability of the function $f(x)$ on $[a, b]$. ■

**Corollary.** *A function which is piecewise continuous on a closed interval is integrable on that interval.*

### 1.8.4. Principal Properties of the Definite Integral

1°. The integral $\int_a^b f(x)\,dx$ was introduced for the case $a<b$. We generalize the concept of the definite integral to the case when the integration limits coincide or when the lower limit is larger than the upper limit. *By definition, we assume that*

$$\int_a^a f(x)\,dx=0, \tag{1}$$

*regarding this formula as a natural extension of the concept of the definite integral to an interval of zero length.*

Also *by definition, we assume that*

$$\int_b^a f(x)\,dx=-\int_a^b f(x)\,dx, \tag{2}$$

*regarding formula* (2) *as a natural extension of the concept of the definite integral to the case when, for $a<b$, the closed interval $[a, b]$ is run over in the direction from $b$ to $a$.* In this case the division points $x_i$ of the interval $[a, b]$ are enumerated in the sequential order from $b$ to $a$ and all the differences $\Delta x_i = x_i - x_{i-1}$ in the integral sum have a negative sign.

2°. *Whatever the numbers $a$, $b$, $c$, the relation*

$$\int_a^b f(x)\,dx=\int_a^c f(x)\,dx+\int_c^b f(x)\,dx \tag{3}$$

*holds true.* (Here and in 1.8.5 we assume that the integrals appearing in the formulas being proved exist.)

*Proof.* We first assume that $a<c<b$. Since the limit of the integral sum $\sigma$ does not depend on how the interval $[a, b]$ is divided, we shall divide $[a, b]$ so that the point $c$ would be a division point. If, for instance, $c = x_m$, then $\sigma$ can be divided into two sums:

$$\sigma=\sum_{i=1}^{n} f(\xi_i)\,\Delta x_i=\sum_{i=1}^{m} f(\xi_i)\,\Delta x_i+\sum_{i=m+1}^{n} f(\xi_i)\,\Delta x_i.$$

Passing to the limit in this relation as $\lambda\to 0$, we get relation (3).

The essence of the property we have proved is that the definite integral over the whole interval is equal to the sum of the integrals over its parts.

The proof for another position of points $a$, $b$, $c$ can be easily reduced to this case. Let, for instance, $a < b < c$. Then, in accordance with what we have proved,

$$\int_a^c f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx,$$

whence, taking (2) into account, we obtain

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx - \int_b^c f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx,$$

i.e. we again arrive at relation (3). ■

3°. *A constant factor can be put before the sign of the definite integral, i.e.*

$$\int_a^b kf(x)\,dx = k\int_a^b f(x)\,dx. \tag{4}$$

*Proof*. Indeed, for any division of the closed interval $[a, b]$ and any choice of the points $\xi_i$

$$\sum_{i=1}^n kf(\xi_i)\,\Delta x_i = k\sum_{i=1}^n f(\xi_i)\,\Delta x_i.$$

Passing to the limit as $\lambda \to 0$, we have

$$\int_a^b kf(x)\,dx = \lim_{\lambda\to 0}\sum_{i=1}^n kf(\xi_i)\,\Delta x_i = \lim_{\lambda\to 0} k\sum_{i=1}^n f(\xi_i)\,\Delta x_i$$

$$= k\lim_{\lambda\to 0}\sum_{i=1}^n f(\xi_i)\,\Delta x_i = k\int_a^b f(x)\,dx,$$

i.e. we have got relation (4). ■

4°. *The definite integral of the algebraic sum of functions is equal to the algebraic sum of their integrals, i.e.*

$$\int_a^b [f(x) \pm g(x)]\,dx = \int_a^b f(x)\,dx \pm \int_a^b g(x)\,dx.$$

*Proof*. Indeed, for any division of the closed interval $[a, b]$ and any choice of points $\xi_i$

$$\sum_{i=1}^n [f(\xi_i) \pm g(\xi_i)]\,\Delta x_i = \sum_{i=1}^n f(\xi_i)\,\Delta x_i \pm \sum_{i=1}^n g(\xi_i)\,\Delta x_i.$$

Since

$$\lim_{\lambda\to 0}\sum_{i=1}^{n} f(\xi_i)\,\Delta x_i = \int_a^b f(x)\,dx \text{ and } \lim_{\lambda\to 0}\sum_{i=1}^{n} g(\xi_i)\,\Delta x_i = \int_a^b g(x)\,dx,$$

it follows that

$$\int_a^b [f(x) \pm g(x)]\,dx = \lim_{\lambda\to 0}\sum_{i=1}^{n} [f(\xi_i) \pm g(\xi_i)]\,\Delta x_i$$

$$= \lim_{\lambda\to 0}\sum_{i=1}^{n} f(\xi_i)\,\Delta x_i \pm \lim_{\lambda\to 0}\sum_{i=1}^{n} g(\xi_i)\,\Delta x_i = \int_a^b f(x)\,dx \pm \int_a^b g(x)\,dx. \blacksquare$$

**Remark.** Property 4° holds true for any finite number of terms.

### 1.8.5. Evaluation of Integrals. Mean Value Formula

**1. Evaluation of integrals** (throughout this section we assume that $a < b$).

1°. *If, throughout the closed interval* $[a, b]$, *the function* $f(x) \geqslant 0$, *then*

$$\int_a^b f(x)\,dx \geqslant 0.$$

*Proof.* Indeed, any integral sum $\sigma = \sum_{i=1}^{n} f(\xi_i)\,\Delta x_i$ for the function $f(x)$ on $[a, b]$ is nonnegative since

$$f(\xi_i) \geqslant 0, \quad \Delta x_i = x_i - x_{i-1} > 0, \quad i = 1, 2, \ldots, n.$$

Passing to the limit in the equality $\sum_{i=1}^{n} f(\xi_i)\,\Delta x_i \geqslant 0$ as $\lambda \to 0$, we obtain

$$\int_a^b f(x)\,dx \geqslant 0. \blacksquare$$

2°. *If, throughout the closed interval* $[a, b]$, $f(x) \leqslant g(x)$, *then*

$$\int_a^b f(x)\,dx \leqslant \int_a^b g(x)\,dx. \tag{1}$$

*Proof.* Applying evaluation 1° to the function $g(x) - f(x) \geqslant 0$, we have

$$\int_a^b [g(x) - f(x)]\,dx \geqslant 0.$$

But, according to Property 4°,

$$\int_a^b [g(x) - f(x)]\,dx = \int_a^b g(x)\,dx - \int_a^b f(x)\,dx \geqslant 0,$$

whence we get inequality (1). ■

3°. *The inequality*

$$\left|\int_a^b f(x)\,dx\right| \leqslant \int_a^b |f(x)|\,dx \tag{2}$$

*holds true for the function $f(x)$ defined on the closed interval $[a, b]$.*

*Proof.* Applying evaluation 2° to the obvious inequalities $-|f(x)| \leqslant f(x) \leqslant |f(x)|$, we obtain

$$-\int_a^b |f(x)|\,dx \leqslant \int_a^b f(x)\,dx \leqslant \int_a^b |f(x)|\,dx,$$

and this is equivalent to inequality (2). ■

**Corollary.** *If, throughout the closed interval $[a, b]$, $|f(x)| \leqslant k$, then*

$$\left|\int_a^b f(x)\,dx\right| \leqslant k(b-a). \tag{3}$$

Indeed, it follows from the inequality $|f(x)| \leqslant k$ and evaluations 2° and 3° that

$$\left|\int_a^b f(x)\,dx\right| \leqslant \int_a^b |f(x)|\,dx \leqslant \int_a^b k\,dx = k\int_a^b dx.$$

From this, noting that

$$\int_a^b dx = \lim_{\lambda\to 0}\sum_{i=1}^{n} 1\cdot\Delta x_i = b-a, \tag{4}$$

we get relation (3).

4°. *If $m$ and $M$ are the least and the greatest value of the function $f(x)$ on the closed interval $[a, b]$ respectively, then*

$$m(b-a) \leqslant \int_a^b f(x)\,dx \leqslant M(b-a). \tag{5}$$

*Proof.* By the hypothesis, for any $x \in [a, b]$ we have

$$m \leqslant f(x) \leqslant M.$$

Applying evaluation 2° to these inequalities, we obtain

$$m \int_a^b dx \leqslant \int_a^b f(x)\, dx \leqslant M \int_a^b dx,$$

whence, taking (4) into account, we obtain (5). ■

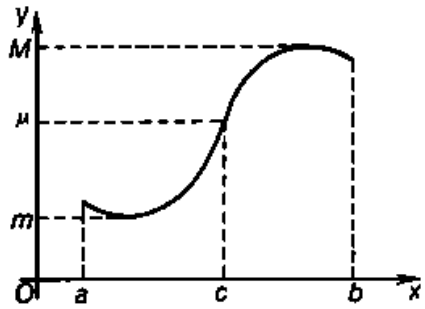



Fig. 98

**2. Mean value formula.**

**Theorem 8.5 (mean value theorem).** *If the function $f(x)$ is continuous on the closed interval $[a, b]$, then on this interval there is a point $c$ such that*

$$\int_a^b f(x)\, dx = f(c)(b-a). \qquad (6)$$

*Proof.* Since $f(x)$ is continuous on $[a, b]$, it follows, in accordance with Weierstrass's second theorem, that there are numbers $m$ and $M$ such that

$$\min_{[a,\, b]} f(x) = m \leqslant f(x) \leqslant M = \max_{[a,\, b]} f(x).$$

From this, by virtue of evaluation 4°, we get

$$m(b-a) \leqslant \int_a^b f(x)\, dx \leqslant M(b-a)$$

and, consequently,

$$m \leqslant \frac{\int_a^b f(x)\, dx}{b-a} \leqslant M.$$

We set

$$\frac{\int_a^b f(x)\, dx}{b-a} = \mu \quad (m \leqslant \mu \leqslant M).$$

Since the number $\mu$ is included between the least and the greatest value of the continuous function $f(x)$ on $[a, b]$ (Fig. 98), it follows, according to Theorem 4.10 on a continuous function running through any intermediate value, that there is a point $c \in [a, b]$ such that

$f(c) = \mu$. Therefore

$$\frac{\int_a^b f(x)\,dx}{b-a} = f(c),$$

and this is equivalent to relation (6). ■

Relation (6) is known as the *mean value formula* and the quantity $f(c)$ as the mean value of the function $f(x)$ on $[a, b]$.

**Remark.** The mean value theorem has the following geometric meaning: for $f(x) \geqslant 0$ the value of the definite integral is equal to the area of a rectangle with altitude $f(c)$ and base $b - a$.

### 1.8.6. An Integral with a Variable Upper Limit

Up till now we considered a definite integral with constant integration limits $a$ and $b$. If we vary the upper limit, for instance, so that we do not go outside the limits of the interval $[a, b]$, then the value of the integral will vary. In other words, an integral with a variable upper limit is a function of its upper limit.

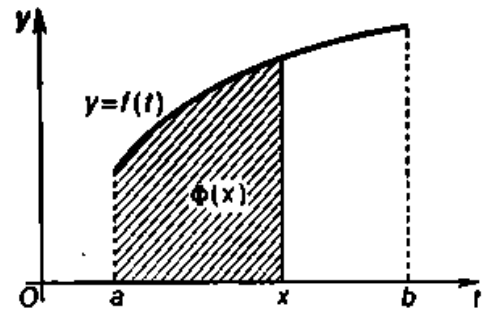


Fig. 99

We shall consider an integral $\int_a^x f(t)\,dt$* $(a \leqslant x \leqslant b)$ with a constant lower limit $a$ and a variable upper limit $x$. The value of this integral is a function of the upper limit $x$. We designate this function as $\Phi(x)$, i.e. set

$$\Phi(x) = \int_a^x f(t)\,dt \tag{1}$$

and term it an *integral with a variable upper limit*. In terms of geometry, the function $\Phi(x)$ is the area of the curvilinear trapezoid hatched in Fig. 99 if $f(x) > 0$.

The following theorem concerns the value of an integral with a variable upper limit.

**Theorem 8.6.** *The derivative of the integral of a continuous function with respect to a variable upper limit exists and is equal to the value*

* For the sake of convenience, we designated the integration variable as $t$ since $x$ is the upper limit of integration.

*of the integrand at the point equal to the upper limit, i.e.*

$$\Phi'(x) = \left(\int_a^x f(t)\,dt\right)' = f(x). \tag{2}$$

*Proof.* We take an arbitrary value $x \in [a, b]$ and assign to it an increment $\Delta x \neq 0$ such that $x + \Delta x \in [a, b]$, i.e. $a \leqslant x + \Delta x \leqslant b$. Then the function $\Phi(x)$ defined by expression (1) assumes a new value

$$\Phi(x + \Delta x) = \int_a^{x+\Delta x} f(t)\,dt.$$

According to Property 2° of the definite integral (see 1.8.4), we have

$$\Phi(x + \Delta x) = \int_a^x f(t)\,dt + \int_x^{x+\Delta x} f(t)\,dt = \Phi(x) + \int_x^{x+\Delta x} f(t)\,dt.$$

From this we find the increment of the function $\Phi(x)$:

$$\Phi(x + \Delta x) - \Phi(x) = \int_x^{x+\Delta x} f(t)\,dt.$$

Applying Theorem 8.5, we get

$$\Phi(x + \Delta x) - \Phi(x) = f(c)\,\Delta x,$$

where $c$ is a number included between the numbers $x$ and $x + \Delta x$. We divide both sides of the relation by $\Delta x$:

$$\frac{\Phi(x + \Delta x) - \Phi(x)}{\Delta x} = f(c).$$

If now $\Delta x \to 0$, then $c \to x$, and then, by virtue of the continuity of the function $f(x)$ on $[a, b]$, $f(c) \to f(x)$. Therefore, passing to the limit as $\Delta x \to 0$ in the last relation, we get

$$\lim_{\Delta x \to 0} \frac{\Phi(x + \Delta x) - \Phi(x)}{\Delta x} = \lim_{\Delta x \to 0} f(c) = \lim_{c \to x} f(c) = f(x)$$

or $\Phi'(x) = f(x)$. ■

We have thus found that any function $f(x)$ continuous on the interval $[a, b]$ has an antiderivative on this interval, the function $\Phi(x)$, which is an integral with a variable upper limit, being an antiderivative of $f(x)$. And since every other antiderivative of the function $f(x)$ can differ from $\Phi(x)$ only by a constant (see Theorem 7.1), we have established the relationship between indefinite

and definite integrals in the form

$$\int f(x)\,\mathrm{d}x = \int_a^x f(t)\,\mathrm{d}t + C,$$

where $C$ is an arbitrary constant.

### 1.8.7. The Newton-Leibniz Formula

It is, as a rule, difficult to calculate definite integrals using the method based on the definition of the integral as a limit of an integral sum. There is a more convenient method of calculating definite integrals which, as will be shown later on, is based on the relationship between indefinite and definite integrals established in 1.8.6.

We have found that the function $f(x)$ continuous on the closed interval $[a, b]$ has antiderivatives, the function

$$\Phi(x) = \int_a^x f(t)\,\mathrm{d}t$$

being one of them.

Let $F(x)$ be any other antiderivative of the function $f(x)$ on the same interval $[a, b]$. Since the antiderivatives $\Phi(x)$ and $F(x)$ differ by a constant, we have a relation

$$\int_a^x f(t)\,\mathrm{d}t = F(x) + C, \quad a \leqslant x \leqslant b,$$

where $C$ is a number. Substituting the value $x = a$ into this relation and using formula (1) from 1.8.4, we have

$$\int_a^a f(t)\,\mathrm{d}t = F(a) + C, \ 0 = F(a) + C, \ C = -F(a),$$

i.e.

$$\int_a^x f(t)\,\mathrm{d}t = F(x) - F(a)$$

for any $x \in [a, b]$. Setting $x = b$, we get the fundamental formula of integral calculus

$$\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a) \tag{1}$$

which is known as the *Newton-Leibniz formula.*

The difference $F(b) - F(a)$ is conventionally written as

$$F(x)\Big|_a^b \text{ or } [F(x)]_a^b$$

and therefore formula (1) assumes the form

$$\int_a^b f(x)\,dx = F(x)\Big|_a^b.$$

It should be emphasized that in formula (1) we can take any antiderivative of $f(x)$ on the interval $[a, b]$ as $F(x)$.

Formula (1) yields a simple method of calculating the definite integral: the definite integral of a continuous function is equal to the difference of the values of any one of its antiderivatives calculated for the upper and the lower limit of integration. This formula opens up wide possibilities for calculating definite integrals since the problem of calculating the definite integral reduces to the problem of calculating an indefinite integral, which has been studied sufficiently thoroughly.

**Examples.**

1. $\int_a^b \sin x\,dx = -\cos x\Big|_a^b = \cos a - \cos b.$

2. $\int_0^2 (3x^2-1)\,dx = [x^3 - x]_0^2 = (2^3-2)-(0^3-0) = 6.$

3. $\int_1^2 \frac{dx}{x} = \ln x\Big|_1^2 = \ln 2 - \ln 1 = \ln 2.$

4. $\int_{-1}^1 \frac{dx}{1+x^2} = \arctan x\Big|_{-1}^1 = \arctan 1 - \arctan(-1) = \frac{\pi}{4} - \left(-\frac{\pi}{4}\right) = \frac{\pi}{4} + \frac{\pi}{4} = \frac{\pi}{2}.$

5. $\int_0^3 \frac{dx}{\sqrt{1+x^2}} = \ln\left(x+\sqrt{1+x^2}\right)\Big|_0^3 = \ln\left(3+\sqrt{10}\right).$

**Remark.** We have derived the Newton-Leibniz formula on the assumption that the integrand function $f(x)$ is continuous. Under certain conditions the Newton-Leibniz formula also holds true for discontinuous functions.

### 1.8.8. A Change of Variable in the Definite Integral

**Theorem 8.7.** *Let $f(x)$ be a continuous function on the closed interval $[a, b]$. Then, if* (1) *the function $x = \varphi(t)$ is differentiable on $[\alpha, \beta]$ and $\varphi'(t)$ is continuous on $[\alpha, \beta]$,* (2) *the interval $[a, b]$ is the set of values of the function $x = \varphi(t)$, and* (3) $\varphi(\alpha) = a$ *and* $\varphi(\beta) = b$ (Fig. 100), *then the formula*

$$\int_a^b f(x)\,dx = \int_\alpha^\beta f[\varphi(t)]\,\varphi'(t)\,dt \qquad (1)$$

*holds true.*

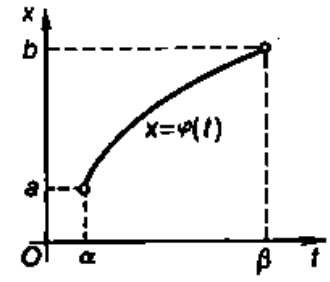


Fig. 100

*Proof.* From the Newton-Leibniz formula we have

$$\int_a^b f(x)\,dx = F(b) - F(a),$$

where $F(x)$ is an antiderivative of the function $f(x)$ on $[a, b]$. On the other hand, we consider a composite function of the variable $t$ on the interval $[\alpha, \beta]$: $\Phi(t) = F[\varphi(t)]$. According to the rule of differentiation of a composite function, we have

$$\Phi'(t) = F'[\varphi(t)]\,\varphi'(t) = f[\varphi(t)]\,\varphi'(t)$$

It follows that the function $\Phi(t)$ is an antiderivative of the function $f[\varphi(t)]\,\varphi'(t)$ continuous on $[\alpha, \beta]$ and therefore, from the Newton-Leibniz formula, we get

$$\int_\alpha^\beta f[\varphi(t)]\,\varphi'(t)\,dt = \Phi(\beta) - \Phi(\alpha) = F[\varphi(\beta)] - F[\varphi(\alpha)]$$

$$= F(b) - F(a) = \int_a^b f(x)\,dx.$$

We have thus proved the validity of formula (1). ■

Formula (1) is known as the *formula for a change of variable* or *for a substitution in a definite integral.*

**Remark 1.** Whereas when calculating an indefinite integral by a change of variable, we must return from the new variable $t$ to the old variable $x$, we may not do that when calculating a definite integral since we must now find a number which, according to the formula we have proved, is equal to the value of each of the integrals in question.

**Example 1.** Calculate $\int_0^1 \sqrt{1-x^2}\,dx$.

*Solution.* We consider the substitution $x = \sin t$, $0 \leqslant t \leqslant \pi/2$. We shall verify whether this substitution is legitimate.

First, the function $f(x) = \sqrt{1 - x^2}$ is continuous on $[0, 1]$; second, the function $x = \sin t$ is differentiable on $[0, \pi/2]$ and $x'_t = \cos t$ is continuous on $[0, \pi/2]$ and, third, when $t$ varies from 0 to $\pi/2$, the function $x = \sin t$ varies from 0 to 1, with $x(0) = 0$ and $x(\pi/2) = 1$. Thus this substitution satisfies all conditions of Theorem 8.7. Using formula (1), we get

$$\int_0^1 \sqrt{1-x^2}\,dx = \int_0^{\pi/2} \sqrt{1-\sin^2 t}\cos t\,dt = \int_0^{\pi/2} \cos^2 t\,dt$$
$$= \frac{1}{2}\int_0^{\pi/2} (1+\cos 2t)\,dt = \frac{\pi}{4}.$$

**Remark 2.** When we use formula (1), we must verify the fulfillment of the conditions enumerated in the theorem. If the conditions are violated, we may get an incorrect result.

**Example 2.** Calculate $\int_0^{\pi} dx$.

*Solution.* We have,

$$\int_0^{\pi} dx = x\Big|_0^{\pi} = \pi.$$

On the other hand

$$\int_0^{\pi} dx = \int_0^{\pi} \frac{dx}{\sin^2 x + \cos^2 x} = \int_0^{\pi} \frac{dx}{\cos^2 x(1+\tan^2 x)}.$$

The substitution $\tan x = t$ formally leads to the result

$$\int_0^{\pi} dx = \int_0^{\pi} \frac{d(\tan x)}{1+\tan^2 x} = \int_0^{0} \frac{dt}{1+t^2} = 0.$$

We have got an incorrect result since $\pi \neq 0$. This happened because the function $t = \tan x$ is discontinuous for $x = \pi/2$ and does not satisfy the conditions of Theorem 8.7.

### 1.8.9. The Formula for Integration by Parts in the Definite Integral

**Theorem 8.8.** *If the functions $u(x)$ and $v(x)$ have continuous derivatives on the closed interval $[a, b]$, then there holds a formula*

$$\int_a^b u\,dv = uv\Big|_a^b - \int_a^b v\,du. \tag{1}$$

*Proof.* Since the function $u(x)\,v(x)$ is an antiderivative of the function $[u(x)\,v(x)]' = u(x)\,v'(x) + v(x)\,u'(x)$, we find from the Newton-Leibniz formula that

$$\int_a^b [u(x)\,v'(x) + v(x)\,u'(x)]\,\mathrm{d}x = [u(x)\,v(x)]_a^b.$$

From this, using Property 4° of the definite integral (see 1.8.4), we obtain

$$\int_a^b u(x)\,v'(x)\,\mathrm{d}x + \int_a^b v(x)\,u'(x)\,\mathrm{d}x = [u(x)\,v(x)]_a^b,$$

or, what is the same,

$$\int_a^b u\,\mathrm{d}v + \int_a^b v\,\mathrm{d}u = uv\,|_a^b,$$

whence follows formula (1). ■

Formula (1) is known as the *formula for integration by parts in the definite integral.*

**Example 1.** Calculate $\int_1^e \ln x\,\mathrm{d}x$.

*Solution.* We set $u = \ln x$, $\mathrm{d}v = \mathrm{d}x$, and then $\mathrm{d}u = \mathrm{d}x/x$, $v = x$ and from formula (1) we find that

$$\int_1^e \ln x\,\mathrm{d}x = x\ln x\,|_1^e - \int_1^e x\,\frac{\mathrm{d}x}{x} = [x\ln x - x]_1^e = 1.$$

**Example 2.** Calculate $\int_1^2 x\mathrm{e}^x\,\mathrm{d}x$.

*Solution.* We set $u = x$, $\mathrm{d}v = \mathrm{e}^x\,\mathrm{d}x$, and then $\mathrm{d}u = \mathrm{d}x$, $v = \mathrm{e}^x$ and from formula (1) we find that

$$\int_1^2 x\mathrm{e}^x\,\mathrm{d}x = x\mathrm{e}^x\,|_1^2 - \int_1^2 \mathrm{e}^x\,\mathrm{d}x = [\mathrm{e}^x(x-1)]_1^2 = \mathrm{e}^2.$$

**Example 3.** Calculate $\int_0^1 \arctan x\,\mathrm{d}x$.

*Solution.* We set $u = \arctan x$, $dv = dx$, and then $du = dx/(1 + x^2)$, $v = x$ and from formula (1) we find that

$$\int_0^1 \arctan x\,dx = x\arctan x\Big|_0^1 - \int_0^1 \frac{x\,dx}{1+x^2}$$

$$= \left[x\arctan x - \frac{1}{2}\ln(1+x^2)\right]_0^1 = \frac{\pi}{4} - \ln\sqrt{2}.$$

### 1.8.10. Some Applications of the Definite Integral in Physics and Geometry

**1. The area of a curvilinear trapezoid.** Assume that a figure given on the $xy$-plane is bounded by the closed interval $[a, b]$ of the $x$-axis, the straight lines $x = a$ and $x = b$ and the graph of the continuous and nonnegative function $y = f(x)$ on $[a, b]$. This is a curvilinear trapezoid whose area $s$ can be calculated by the formula

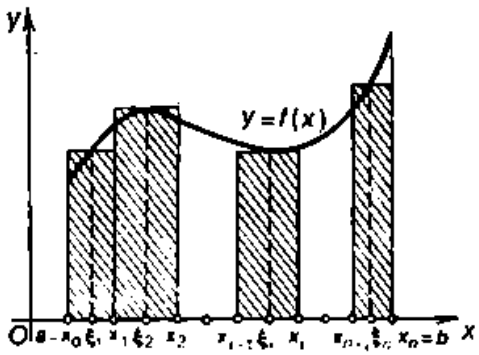


Fig. 101

$$s = \int_a^b f(x)\,dx. \qquad (1)$$

*Proof.* We divide arbitrarily the closed interval $[a, b]$ into $n$ parts by the points $a = x_0 < x_1 < x_2 < \ldots < x_{i-1} < x_i < \ldots < x_n = b$, choose an arbitrary point $\xi_i$ $(x_{i-1} \leqslant \xi_i \leqslant x_i)$ on each subinterval $[x_{i-1}, x_i]$ and consider the stepped figure shown in Fig. 101. The area $s$ of the curvilinear trapezoid is approximately equal to the area of the stepped figure:

$$s \cong \sum_{i=1}^{n} f(\xi_i)\,\Delta x_i,$$

where $\Delta x_i = x_i - x_{i-1}$. It is natural to assume that, as $\lambda = \max\limits_{1\leqslant i\leqslant n}\{\Delta x_i\} \to 0$, the area of the stepped figure tends to the area of the curvilinear trapezoid. On the other hand, the area of the stepped figure is the integral sum for integral (1). Since the function $f(x)$ is continuous on $[a, b]$, the limit of this sum exists for $\lambda = \max\limits_{1\leqslant i\leqslant n}\{\Delta x_i\} \to 0$ and is equal to the integral of the function $f(x)$ over the interval $[a, b]$. Consequently, the area $s$ of the curvilinear trapezoid is also numerically equal to the definite integral of the

function $f(x)$ over $[a, b]$, i.e.

$$s = \lim_{\lambda \to 0} \sum_{i=1}^{n} f(\xi_i)\,\Delta x_i = \int_a^b f(x)\,dx. \quad \blacksquare$$

Thus *a definite integral of a nonnegative continuous function $f(x)$ taken over $[a, b]$ is numerically equal to the area of a curvilinear trapezoid with the base $[a, b]$, bounded by the graph of the function $y = f(x)$ from above.* This is the geometric meaning of the definite integral.

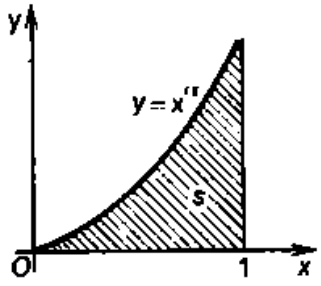



Fig. 102

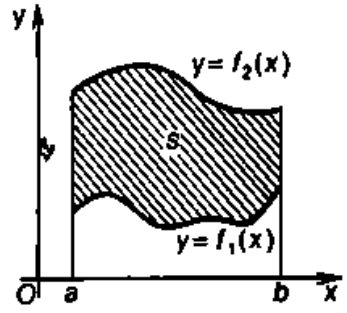



Fig. 103

**Example 1.** Find the area of the figure bounded by the graph of the function $y = x^\alpha$, $\alpha > 0$, the straight line $x = 1$ and the $x$-axis (Fig. 102).

*Solution.* From formula (1) we find that

$$s = \int_0^1 x^\alpha\,dx = \frac{x^{\alpha+1}}{\alpha+1}\Big|_0^1 = \frac{1}{\alpha+1}.$$

If $\alpha = 1$, then $s = 1/2$; if $\alpha = 2$, then $s = 1/3$, and so on.

Assume that the figure is bounded from below and from above by the graphs of the functions $y = f_1(x)$ and $y = f_2(x)$, $f_1(x) \leqslant f_2(x)$, $a \leqslant x \leqslant b$ (Fig. 103), where $f_1(x)$ and $f_2(x)$ are two continuous functions. If both functions are nonnegative, then the area $s$ of the given figure is equal to the difference of the areas of the curvilinear trapezoids bounded from above by the graphs of the functions $y = f_2(x)$ and $y = f_1(x)$ respectively. Consequently,

$$s = \int_a^b f_2(x)\,dx - \int_a^b f_1(x)\,dx = \int_a^b [f_2(x) - f_1(x)]\,dx. \tag{2}$$

Note that formula (2) is also valid when $f_1(x)$ and $f_2(x)$ are not nonnegative. Indeed, by virtue of their boundedness, there is a number $h > 0$ such that the functions $f_1^*(x) = f_1(x) + h$ and $f_2^*(x) =$

$f_2(x) + h$ are nonnegative and there holds an obvious equality

$$\int_a^b [f_2^*(x) - f_1^*(x)]\,dx = \int_a^b [f_2(x) - f_1(x)]\,dx.$$

**Example 2.** Calculate the area of the figure bounded by the graphs of the functions $y = f_1(x) = x$ and $y = f_2(x) = 2 - x^2$ (Fig. 104).

*Solution.* We find the abscissas of the points of intersection of the straight line $y = x$ and the parabola $y = 2 - x^2$. Solving the simultaneous equations

$$\begin{cases} y = x, \\ y = 2 - x^2, \end{cases}$$

we find that $x_1 = -2$ and $x_2 = 1$. These are the limits of integration. According to formula (2), the required area of the figure is

$$s = \int_{-2}^{1} [f_2(x) - f_1(x)]\,dx = \int_{-2}^{1} [(2 - x^2) - x]\,dx$$

$$= \left[2x - \frac{x^3}{2} - \frac{x^2}{2}\right]_{-2}^{1} = \frac{9}{2}.$$

**Remark.** To calculate the area of a curvilinear trapezoid in the case when the upper bound is given by the parametric equations $x = \varphi(t)$ and $y = \psi(t)$, $\alpha \leqslant t \leqslant \beta$, with $\varphi(\alpha) = a$ and $\varphi(\beta) = b$, we must make a change of variable in formula (1) setting $x = \varphi(t)$, $dx = \varphi'(t)\,dt$. Then we get

$$s = \int_\alpha^\beta \psi(t)\,\varphi'(t)\,dt.$$

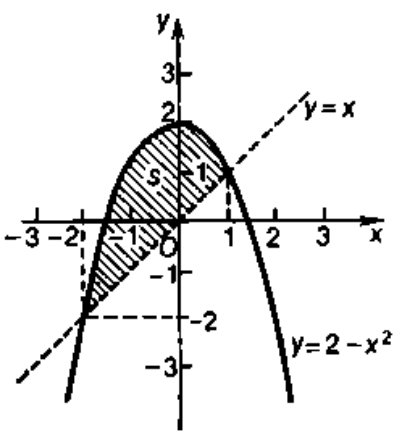



Fig. 104

**Example 3.** Calculate the area of the figure bounded by the ellipse

$$\begin{cases} x = a\cos t, \\ y = b\sin t, \end{cases} \quad 0 \leqslant t \leqslant 2\pi.$$

*Solution.* An ellipse is symmetric with respect to the axes of coordinates and it is therefore sufficient to calculate the area of the part of the figure which is in the first quadrant (Fig. 105). Consequently, the required area is

$$s = 4\int_{\pi/2}^{0} b\sin t(a\cos t)'\,dt = 4ab\int_0^{\pi/2} \sin^2 t\,dt$$

$$= 2ab\int_0^{\pi/2} (1 - \cos 2t)\,dt = 2ab\left[t - \frac{1}{2}\sin 2t\right]_0^{\pi/2} = \pi ab.$$

In particular, if $a = b = R$, then we get the familiar formula for the area of a circle $\pi R^2$.

**2. The area of a curvilinear sector.** Assume that the curve $AB$ is defined in the polar coordinates by the equation $\rho = \rho(\varphi)$, $\alpha \leqslant \varphi \leqslant \beta$, the function $\rho(\varphi)$ being continuous and nonnegative on the closed interval $[\alpha, \beta]$. We call the plane figure bounded by the curve $AB$ and two rays which make angles $\alpha$ and $\beta$ with the polar axis a *curvilinear sector* (Fig. 106). The area $s$ of the curvilinear sector

$$s = \frac{1}{2} \int_{\alpha}^{\beta} \rho^2(\varphi)\, d\varphi. \tag{3}$$

*Proof.* We divide arbitrarily the closed interval $[\alpha, \beta]$ into $n$ parts by the points $\alpha = \varphi_0 < \varphi_1 < \varphi_2 < \ldots < \varphi_{i-1} < \varphi_i < \ldots < \varphi_n = \beta$, choose an arbitrary point $\xi_i$ $(\varphi_{i-1} \leqslant \xi_i \leqslant \varphi_i)$

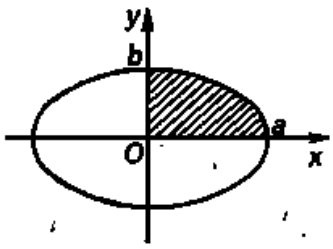



Fig. 105

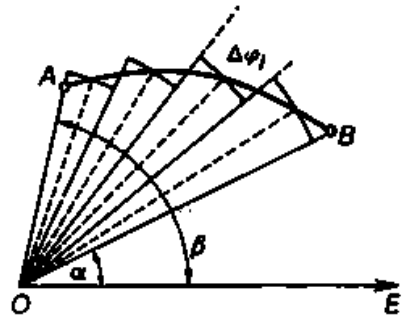



Fig. 106

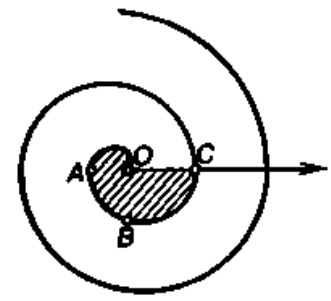



Fig. 107

on each of the subintervals $[\varphi_{i-1}, \varphi_i]$ and construct circular sectors with radii $\rho(\xi_i)$. We get a fan-shaped figure whose area is approximately equal to the area $s$ of the curvilinear sector, i.e.

$$s \cong \frac{1}{2} \sum_{i=1}^{n} \rho^2(\xi_i)\, \Delta\varphi_i,$$

where $\Delta\varphi_i = \varphi_i - \varphi_{i-1}$. On the other hand, the area of the fan-shaped figure is the integral sum for integral (3). Since the function $\rho^2(\varphi)$ is continuous on the interval $[\alpha, \beta]$, the limit of this sum exists for $\lambda = \max\limits_{1 \leqslant i \leqslant n} \{\Delta\varphi_i\} \to 0$ and is equal to integral (3). Consequently, the area of the curvilinear sector is also numerically equal to this definite integral:

$$s = \frac{1}{2} \lim_{\lambda \to 0} \sum_{i=1}^{n} \rho^2(\xi_i)\, \Delta\varphi_i = \frac{1}{2} \int_{\alpha}^{\beta} \rho^2(\varphi)\, d\varphi. \ \blacksquare$$

**Example 4.** Calculate the area of the figure bounded by the polar axis and the first turn of the spiral of Archimedes: $\rho = a\varphi$, where $a$ is a positive number (Fig. 107).

*Solution.* When $\varphi$ varies from 0 to $2\pi$, the polar radius describes a curve which bounds the curvilinear sector $OABC$. Therefore from formula (3) we have

$$s_{OABC} = \frac{a^2}{2}\int_0^{2\pi} \varphi^2\,d\varphi = \frac{a^2}{2}\frac{\varphi^3}{3}\Big|_0^{2\pi} = \frac{a^2}{2}\frac{8\pi^3}{3} = \frac{4}{3}\pi^3 a^2.$$

The distance from the point $C$ to the pole is $\rho = 2\pi a$. Therefore a circle of radius $OC$ has an area $\pi \cdot OC^2 = 4\pi^3 a^2 = 3\cdot\frac{4}{3}\pi^3 a^2 = 3s_{OABC}$, i.e. the area of the figure bounded by the polar axis and the first turn of the spiral of Archimedes is equal to 1/3 of the area of the circle with radius equal to the largest of the polar radii of the turn. It was Archimedes who arrived at this conclusion.

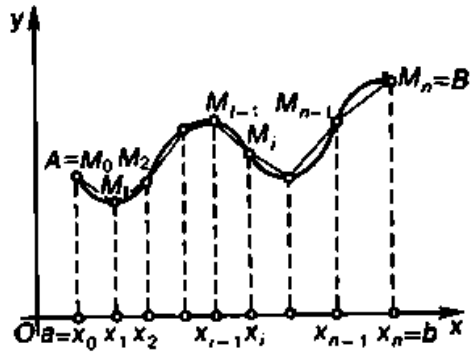



Fig. 108

**3. The length of the arc of a curve.** Assume that the plane curve $AB$ is defined by the equation $y = f(x)$, $a \leqslant x \leqslant b$, where $f(x)$ is a continuous function on the closed interval $[a, b]$. We divide the curve $AB$ into $n$ arbitrary parts by the points $A = M_0$, $M_1$, $M_2$, ..., $M_{i-1}$, $M_i$, ..., $M_n = B$ in the direction from $A$ to $B$. Connecting the neighbouring points by chords, we get a polygonal line inscribed into the curve $AB$, whose length we designate as $P$ (Fig. 108). We designate the length of one segment $M_{i-1}M_i$ of the polygonal line as $l_i$ and the length of the largest of its segments as $\mu$, i.e. $\mu = \max\limits_{1\leqslant i\leqslant n} \{l_i\}$.

**Definition.** *The number $L$ is the limit of the lengths $P$ of the polygonal lines as $\mu \to 0$ ($L = \lim\limits_{\mu\to 0} P$) if for any $\varepsilon > 0$ there is $\delta > 0$ such that for every polygonal line in which $\mu < \delta$ the inequality*

$$|L - P| < \varepsilon$$

*holds true.*

If there is a limit $L$ of the lengths $P$ of the polygonal lines inscribed into the curve as $\mu \to 0$, then this limit is the *length of the arc $AB$*.

If the function $f(x)$ is continuous together with $f'(x)$ on the closed interval $[a, b]$, then the length $L$ of the arc $AB$ is expressed by the formula

$$L = \int_a^b \sqrt{1 + f'^2(x)}\,dx. \tag{4}$$

*Proof.* We designate the coordinates of the point $M_i$ as $(x_i, f(x_i))$ so that for the abscissas of these points we get $a = x_0 < x_1 < x_2 < \ldots < x_{i-1} < x_i < \ldots < x_n = b$. Then the lenght $l_i$ of one segment of the polygonal line is

$$l_i = \sqrt{(x_i - x_{i-1})^2 + [f(x_i) - f(x_{i-1})]^2}.$$

From Lagrange's formula we have

$$f(x_i) - f(x_{i-1}) = f'(\xi_i)(x_i - x_{i-1}), \quad x_{i-1} \leqslant \xi_i \leqslant x_i.$$

Consequently, $l_i = \sqrt{1 + f'^2(\xi_i)}\,\Delta x_i, \quad \Delta x_i = x_i - x_{i-1}$.

Thus the length of the whole polygonal line is

$$P = \sum_{i=1}^{n} l_i = \sum_{i=1}^{n} \sqrt{1 + f'^2(\xi_i)}\,\Delta x_i.$$

The right-hand side of the relation is the integral sum for integral (4). The function $\sqrt{1 + f'^2(x)}$ is continuous on $[a, b]$ and therefore

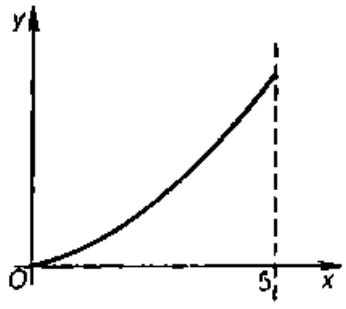



Fig. 109

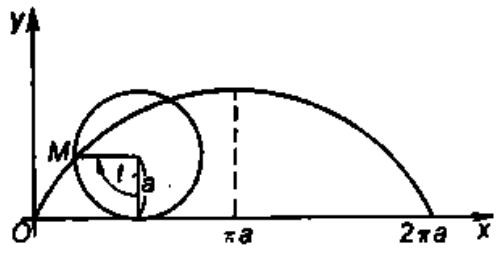



Fig. 110

the limit of this sum exists as $\lambda = \max\limits_{1 \leqslant i \leqslant n} \{\Delta x_i\} \to 0$ and is equal to the definite integral (4). Since $\lambda \leqslant \mu$*, it follows that $\lambda \to 0$ as $\mu \to 0$. Consequently,

$$L = \lim_{\mu \to 0} P = \lim_{\lambda \to 0} \sum_{i=1}^{n} \sqrt{1 + f'^2(\xi_i)}\,\Delta x_i = \int_a^b \sqrt{1 + f'^2(x)}\,dx. \ \blacksquare$$

**Example 5.** Calculate the length of the arc of the semicubical parabola $y = x^{3/2}$ if $0 \leqslant x \leqslant 5$ (Fig. 109).

*Solution.* From the equation $y = x^{3/2}$ we find that $y' = (3/2)\,x^{1/2}$. Consequently, from formula (4) we get

$$L = \int_0^5 \sqrt{1 + y'^2}\,dx = \int_0^5 \sqrt{1 + \frac{9x}{4}}\,dx = \frac{8}{27}\left(1 + \frac{9x}{4}\right)^{3/2}\Big|_0^5 = \frac{335}{27}.$$

* $l_i = \sqrt{(\Delta x_i)^2 + (\Delta y_i)^2}$, whence $|\Delta x_i| \leqslant l_i$.

**Remark 1.** To calculate the length of the arc in the case when the curve $AB$ is defined by the parametric equations $x = \varphi(t)$ and $y = \psi(t)$, $\alpha \leqslant t \leqslant \beta$, where $\alpha$ and $\beta$ are the values of the parameter $t$ corresponding to the values $x = a$ and $x = b$, i.e. $a = \varphi(\alpha)$ and $b = \varphi(\beta)$, we must make a change of variable in the formula $L = \int_a^b \sqrt{1+y'^2(x)}\,dx$, setting $x=\varphi(t)$, $dx=\varphi'(t)\,dt$. Then we obtain

$$L=\int_a^b \sqrt{1+y'^2(x)}\,dx=\int_\alpha^\beta \sqrt{1+\left(\frac{\psi'(t)}{\varphi'(t)}\right)^2}\,\varphi'(t)\,dt$$

$$=\int_\alpha^\beta \sqrt{\varphi'^2(t)+\psi'^2(t)}\,dt. \tag{5}$$

**Example 6.** Calculate the length of one arc of the cycloid*: $x = a(t - \sin t)$, $y = a(1 - \cos t)$, $0 \leqslant t \leqslant 2\pi$ (Fig. 110).

*Solution.* From the equations of a cycloid we find $\varphi'(t) = a(1 - \cos t)$ and $\psi'(t) = a \sin t$. When $x$ runs over the interval $[0, 2\pi a]$, the parameter $t$ runs over the interval $[0, 2\pi]$. Consequently, the required length of the arc is

$$L=\int_0^{2\pi a} \sqrt{1+y'^2(x)}\,dx=\int_0^{2\pi} \sqrt{\varphi'^2(t)+\psi'^2(t)}\,dt$$

$$=\int_0^{2\pi} a\sqrt{(1-\cos t)^2+\sin^2 t}\,dt=2a\int_0^{2\pi}\sin\frac{t}{2}\,dt=-4a\cos\frac{t}{2}\Big|_0^{2\pi}=8a.$$

**Remark 2.** To calculate the length of the arc in the case when the curve $AB$ is defined in polar coordinates by the equation $\rho = \rho(\varphi)$, $\alpha \leqslant \varphi \leqslant \beta$, where $\rho(\varphi)$ has a continuous derivative $\rho'(\varphi)$ on the interval $[\alpha, \beta]$ and the points $A$ and $B$ are associated with the values of $\varphi$ equal to $\alpha$ and $\beta$, we must pass from the polar coordinates (see formula (1) in 1.3.3) to the rectangular coordinates. Then we get a parametric representation of the curve $AB$ by the equations $x = \rho(\varphi)\cos\varphi$ and $y = \rho(\varphi)\sin\varphi$, $\alpha \leqslant \varphi \leqslant \beta$ ($\varphi$ is a parameter). Since

$$x'(\varphi) = \rho'(\varphi)\cos\varphi - \rho(\varphi)\sin\varphi,$$

$$y'(\varphi) = \rho'(\varphi)\sin\varphi + \rho(\varphi)\cos\varphi,$$

* A cycloid is a plane curve which is described by a point $M$ of a circle of radius $a$ rolling without slipping along a straight line.

formula (5) assumes the form

$$L=\int_{\alpha}^{\beta}\sqrt{\rho(\varphi)+\rho'^2(\varphi)}\,d\varphi. \tag{6}$$

**Example 7.** Calculate the length of the first turn of the spiral of Archimedes: $\rho = a\varphi$ (see Fig. 107).

*Solution.* The first turn of the spiral results from the variation of the polar angle $\varphi$ from 0 to $2\pi$. Therefore, from formula (6), the required length of the arc is

$$L=\int_0^{2\pi}\sqrt{a^2\varphi^2+a^2}\,d\varphi=a\int_0^{2\pi}\sqrt{\varphi^2+1}\,d\varphi$$

$$=a\left[\pi\sqrt{4\pi^2+1}+\frac{1}{2}\ln\left(2\pi+\sqrt{4\pi^2+1}\right)\right].$$

**4. The volume of a body of revolution.** Let the function $f(x)$ be continuous and nonnegative on the closed interval $[a, b]$. Then the body which results from the rotation of a curvilinear trapezoid bounded from above by the graph of the function $y = f(x)$ about the $x$-axis has a volume

$$v=\pi\int_a^b f^2(x)\,dx. \tag{7}$$

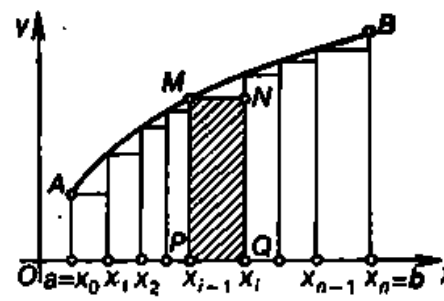



Fig. 111

*Proof.* We divide arbitrarily the closed interval $[a, b]$ into $n$ parts by the points $a = x_0 < x_1 < x_2 < \ldots < x_{i-1} < x_i < \ldots < x_n = b$. Then we construct a rectangle on each subinterval $[x_{i-1}, x_i]$ (Fig. 111). Rotating about the $x$-axis, each rectangle describes a cylinder. We find the volume of the $i$th cylinder resulting from the rotation of the rectangle $PMNQ$:

$$v_i=\pi f^2(x_{i-1})\,\Delta x_i,$$

where $\Delta x_i = x_i - x_{i-1}$.

The sum of the volumes of all the $n$ cylinders is approximately equal to the volume of the given body of revolution:

$$v\cong\sum_{i=1}^{n}\pi f^2(x_{i-1})\,\Delta x_i.$$

On the other hand, this is the integral sum for integral (7). Since the function $f^2(x)$ is continuous on $[a, b]$, the limit of this sum exists as $\lambda = \max\limits_{1\leqslant i\leqslant n}\{\Delta x_i\}\to 0$ and is equal to the definite integral (7).

Thus

$$v = \lim_{\lambda \to 0} \sum_{i=1}^{n} \pi f^2(x_{i-1})\,\Delta x_i = \pi \int_a^b f^2(x)\,dx. \quad \blacksquare$$

**Example 8.** Calculate the volume of a torus*.

*Solution.* Let a circle rotate about the $x$-axis (Fig. 112). We can represent the volume of the torus as the difference of the volumes of bodies resulting from the rotation of the curvilinear trapezoids $ABCDE$ and $ABLDE$ about the $x$-axis.

The equation of the circumference $LBCD$ has the form

$$x^2 + (y - b)^2 = a^2,$$

the equation of the curve $BCD$ being

$$y = y_1(x) = b + \sqrt{a^2 - x^2},$$

and the equation of the curve $BLD$ being

$$y = y_2(x) = b - \sqrt{a^2 - x^2}.$$

Using formula (7), we get an expression

$$v = 2\pi \int_0^a y_1^2\,dx - 2\pi \int_0^a y_2^2\,dx = 2\pi \int_0^a (y_1^2 - y_2^2)\,dx$$

$$= 2\pi \int_0^a [(b + \sqrt{a^2 - x^2})^2 - (b - \sqrt{a^2 - x^2})^2]\,dx$$

$$= 8\pi b \int_0^a \sqrt{a^2 - x^2}\,dx = 2\pi^2 a^2 b$$

for the volume $v$ of the torus.

**5. The area of a surface of revolution.** Let the function $f(x)$ be continuous and nonnegative together with its first derivative on the closed interval $[a, b]$. Then the surface resulting from the rotation of the graph of this function about the $x$-axis has the area $P$ which can be calculated by the formula

$$P = 2\pi \int_a^b f(x) \sqrt{1 + f'^2(x)}\,dx. \tag{8}$$

*Proof.* We divide arbitrarily the closed interval $[a, b]$ into $n$ parts by the points $a = x_0 < x_1 < x_2 < \ldots < x_{i-1} < x_i < \ldots < x_n =$

* A torus is a body resulting from the rotation of a circle of radius $a$ about the axis which lies in its plane at the distance $b$ from the centre of the circle ($b \geqslant a$).

*b*. Let $A_0, A_1, A_2, \ldots, A_{i-1}, A_i, \ldots, A_n$ be the corresponding points of the graph of the function $f(x)$. We construct a polygonal line $A_0, A_1, A_2, \ldots, A_n$ (Fig. 113). When we rotate this polygonal line about the $x$-axis, we get a surface formed by the lateral surfaces of truncated cones (cylinders). The area of the lateral surface

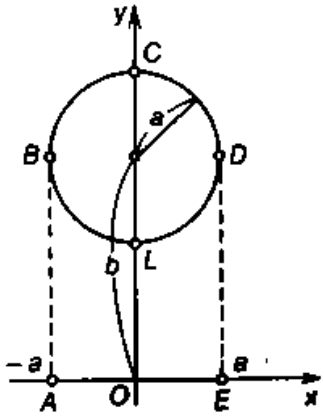



Fig. 112

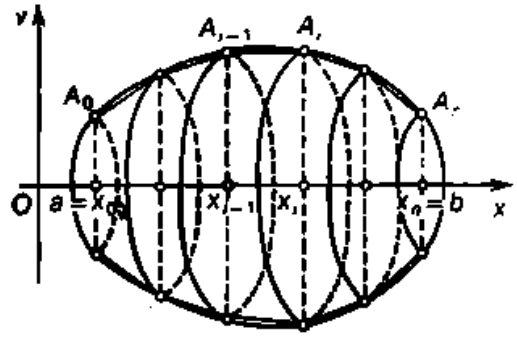



Fig. 113

of a truncated cone (cylinder) resulting from the rotation of the $i$th segment of the polygonal line is $2\pi\frac{f(x_{i-1})+f(x_i)}{2}\,l_i$, where $l_i$ is the length of the chord $A_{i-1}A_i$, i.e.

$$l_i=\sqrt{(x_i-x_{i-1})^2+[f(x_i)-f(x_{i-1})]^2}.$$

By Lagrange's formula we have

$$f(x_i)-f(x_{i-1})=f'(\xi_i)(x_i-x_{i-1}),\quad x_{i-1}\leqslant\xi_i\leqslant x_i.$$

Setting $x_i-x_{i-1}=\Delta x_i$, we obtain

$$l_i=\sqrt{1+f'^2(\xi_i)}\,\Delta x_i.$$

Thus the area $P$ of the surface of revolution is approximately equal to the area of the surface resulting from the rotation of the polygonal line

$$P\cong\sum_{i=1}^{n}2\pi\frac{f(x_{i-1})+f(x_i)}{2}\sqrt{1+f'^2(\xi_i)}\,\Delta x_i.$$

We represent this sum as two sums

$$P\cong 2\pi\sum_{i=1}^{n}f(\xi_i)\sqrt{1+f'^2(\xi_i)}\,\Delta x_i$$

$$+\pi\left\{\sum_{i=1}^{n}[(f(x_{i-1})-f(\xi_i))+(f(x_i)-f(\xi_i))]\sqrt{1+f'^2(\xi_i)}\,\Delta x_i\right\}. \qquad (9)$$

The first sum on the right-hand side of the last relation is the integral sum for integral (8) and, as $\lambda = \max\limits_{1 \leqslant i \leqslant n} \{\Delta x_i\} \to 0$ by virtue of the continuity of the function $f(x)\sqrt{1+f'^2(x)}$, has this integral as its limit. We shall show that the expression in braces on the right-hand side of relation (9) has a zero limit as $\lambda \to 0$. Indeed, since the function $f(x)$ is uniformly continuous on $[a, b]$, we have, according to Cantor's theorem, that for any $\varepsilon > 0$ there is $\delta > 0$ such that, for $\lambda < \delta$, the inequalities $|f(x_{i-1}) - f(\xi_i)| < \varepsilon$ and $|f(x_i) - f(\xi_i)| < \varepsilon$ hold true. If we designate as $M$ the maximum value of the function $\sqrt{1+f'^2(x)}$ on the interval $[a, b]$, then, for $\lambda < \delta$, the expression in braces can be evaluated as follows:

$$\left|\left\{\sum_{i=1}^{n} [(f(x_{i-1}) - f(\xi_i)) + (f(x_i) - f(\xi_i))]\sqrt{1+f'^2(\xi_i)}\,\Delta x_i\right\}\right|$$

$$< 2M\varepsilon \sum_{i=1}^{n} \Delta x_i = 2M(b-a)\,\varepsilon.$$

Since $\varepsilon$ is arbitrarily small, it follows that the limit of this expression is zero as $\lambda \to 0$.

Thus, passing to the limit in relation (9) as $\lambda \to 0$, we have

$$P = 2\pi \int_a^b f(x)\sqrt{1+f'^2(x)}\,dx,$$

i.e. we have obtained the required formula (8). ■

**Remark.** If a surface results from the rotation, about the $x$-axis, of a curve $AB$ represented by the parametric equations $x = \varphi(t)$ and $y = \psi(t)$, $\alpha \leqslant t \leqslant \beta$, where $\psi(t) \geqslant 0$ and $\varphi(t)$ varies from $a$ to $b$ when $t$ varies from $\alpha$ to $\beta$, $\varphi(\alpha) = a$ and $\varphi(\beta) = b$, then, making a change of variable $x = \varphi(t)$ in integral (8), we obtain

$$P = 2\pi \int_\alpha^\beta \psi(t)\sqrt{\varphi'^2(t) + \psi'^2(t)}\,dt. \tag{10}$$

Finally, if the curve is defined by an equation in polar coordinates, i.e. $\rho = \rho(\varphi)$, $\alpha \leqslant \varphi \leqslant \beta$, where $\rho(\varphi)$ has a continuous derivative on $[\alpha, \beta]$, then this case, as was pointed out in Item 3, reduces to the parametric representation of the curve $x = \rho(\varphi)\cos\varphi$, $y = \rho(\varphi)\sin\varphi$, $\alpha \leqslant \varphi \leqslant \beta$, and formula (10) assumes the form

$$P = 2\pi \int_\alpha^\beta \rho(\varphi)\sin\varphi\sqrt{\rho^2(\varphi) + \rho'^2(\varphi)}\,d\varphi.$$

**Example 9.** Calculate the area $P$ of the surface of a spherical zone resulting from the rotation of a semicircle $f(x) = \sqrt{R^2 - x^2}$, $-R < a \leqslant x \leqslant b < R$, about the $x$-axis.

*Solution.* From formula (8) we get

$$P = 2\pi \int_a^b \sqrt{R^2 - x^2} \sqrt{1 + \frac{x^2}{R^2 - x^2}}\, dx = 2\pi \int_a^b R\, dx$$

$$= 2\pi R(b - a) = 2\pi R h,$$

where $h$ is the height of the zone.

**Example 10.** Calculate the area $P$ of the surface resulting from the rotation of one arc of the cycloid: $x = a(t - \sin t)$, $y = a(1 - \cos t)$, $0 \leqslant t \leqslant 2\pi$, about the $x$-axis.

*Solution.* From formula (10) we have

$$P = 2\pi \int_0^{2\pi} a(1 - \cos t) \sqrt{(a \sin t)^2 + [a(1 - \cos t)]^2}\, dt$$

$$= 2\sqrt{2}\, \pi a^2 \int_0^{2\pi} (1 - \cos t)^{3/2}\, dt = \frac{64}{3} \pi a^2.$$

**6. The work of a variable force.** From the problems connected with the geometric application of the definite integral we have considered it follows that one and the same method of calculation is used to solve them: the approximate value of the required quantity is represented as the integral sum and then, as a result of a limiting process, the exact value is obtained in the form of an integral. This method is used to solve a variety of other problems in mechanics, physics and engineering. We shall calculate the work of a variable force as an example.

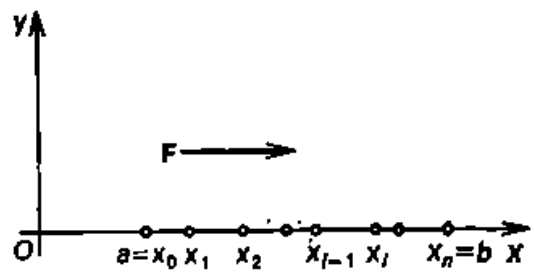


Fig. 114

Assume that a particle moves under the action of a force $F$ directed along the $x$-axis and having a variable value dependent on $x$. We have to find the work $A$ done by the force $F$ in displacing the particle along the $x$-axis from the point $x = a$ to the point $x = b$ $(a < b)$. We assume the function $F(x)$ to be continuous on the closed interval $[a, b]$ (Fig. 114).

We divide arbitrarily the interval $[a, b]$ into $n$ parts by the points $a = x_0 < x_1 < x_2 < \ldots < x_{i-1} < x_i < \ldots < x_n = b$ and choose a point $\xi_i$ on each subinterval $[x_{i-1}, x_i]$. The force acting on the particle on the subinterval $[x_{i-1}, x_i]$ varies from point to point. But

if the length of the subinterval is small, the value of the force at the points of the subinterval $[x_{i-1}, x_i]$ differs but little from its value at any point $\xi_i \in [x_{i-1}, x_i]$ since $F(x)$ is continuous. We can therefore assume the work $A_i$ done by the force $F$ on $[x_{i-1}, x_i]$ to be approximately equal to the work done by the constant force $F(\xi_i)$ on the same subinterval, i.e.

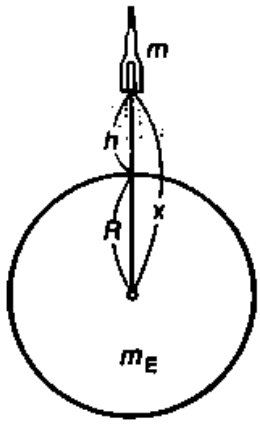


Fig. 115

$$A_i \cong F(\xi_i)\,\Delta x_i.$$

Reasoning by analogy for every subinterval, we get the approximate value for the work $A$ of the force $F$ on the whole interval:

$$A \cong \sum_{i=1}^{n} F(\xi_i)\,\Delta x_i.$$

On the other hand, the sum on the right-hand side of the relation is the integral sum for the function $F(x)$. Since the function $F(x)$ is continuous on $[a, b]$, the limit of this sum exists as $\lambda = \max\limits_{1 \leqslant i \leqslant n} \{\Delta x_i\} \to 0$ and is equal to the definite integral of the function $F(x)$ on the interval $[a, b]$. Thus

$$A = \lim_{\lambda \to 0} \sum_{i=1}^{n} F(\xi_i)\,\Delta x_i = \int_a^b F(x)\,dx. \tag{11}$$

**Example 11.** Find the work $A$ needed to launch a body of mass $m$ from the surface of the Earth vertically upwards to the height $h$ (Fig. 115).

*Solution.* We denote the force of attraction of the body by the Earth by $F$. Let $m_E$ be the mass of the Earth. According to Newton's law

$$F = G\frac{mm_E}{x^2},$$

where $x$ is the distance from the body to the centre of the Earth. We set $Gmm_E = K$ and get $F(x) = K/x^2$, $R \leqslant x \leqslant h + R$, where $R$ is the radius of the Earth. For $x = R$ the force $F(R)$ is equal to the weight of the body $P = mg$, i.e. $K/R^2 = P$, whence $K = PR^2$ and $F(x) = PR^2/x^2$. Thus, from formula (11), we obtain

$$A = \int_R^{R+h} F(x)\,dx = PR^2 \int_R^{R+h} \frac{dx}{x^2} = -PR^2 \frac{1}{x}\Big|_R^{R+h} = \frac{PRh}{R+h}.$$

### 1.8.11. Improper Integrals

Introducing the definite integral as the limit of integral sums, we assumed that the integration interval is finite and the integrand function is bounded on that interval. If at least one of these conditions is not fulfilled, the definition of the definite integral given above loses sense. Thus, in the case when the integration interval is infinite, we cannot divide it into $n$ parts of a finite length, and in the case when the function is unbounded, the integral sum does not have a finite limit. However, we can generalize the concept of the definite integral to these cases as well. The concept of an improper integral resulted precisely from this generalization.

**1. Improper integrals with infinite limits of integration.**

**Definition.** *Assume that the function $f(x)$ is defined on the half-open interval $[a, +\infty)$ and integrable over any closed interval $[a, R]$, i.e. there is a definite integral $\int_a^R f(x)\,dx$ for any $R > a$. If there is a finite limit*

$$\lim_{R\to+\infty} \int_a^R f(x)\,dx, \tag{1}$$

*then it is called an improper integral of the first kind and is designated as*

$$\int_a^{+\infty} f(x)\,dx. \tag{2}$$

Thus, by definition,

$$\int_a^{+\infty} f(x)\,dx = \lim_{R\to+\infty} \int_a^R f(x)\,dx.$$

In this case we say that integral (2) exists or converges. Now if limit (1) does not exist or is infinite, we say that integral (2) does not exist or diverges.

We can introduce an improper integral over the interval $(-\infty, b]$ by analogy with integral (2):

$$\int_{-\infty}^{b} f(x)\,dx = \lim_{R\to-\infty} \int_R^b f(x)\,dx. \tag{3}$$

Finally, we can define an improper integral with two infinite limits as the sum of integrals of forms (2) and (3), i.e.

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_{-\infty}^{c} f(x)\,dx + \int_c^{+\infty} f(x)\,dx, \tag{4}$$

where $c$ is any number, provided that both integrals exist from the right.

We shall establish the geometric meaning of an improper integral of the first kind. Let $f(x) \geqslant 0$. Then the definite integral $\int_a^R f(x)\,dx$ expresses the area of the domain bounded by the graph of the function $f(x)$ from above, by the $x$-axis from below, by the straight line $x = a$

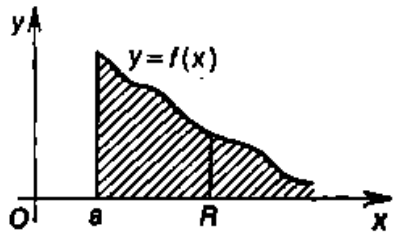



Fig. 116

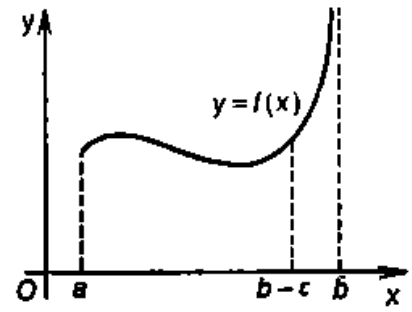



Fig. 117

from the left and by the straight line $x = R$ from the right. It is natural to assume that the improper integral $\int_a^{+\infty} f(x)\,dx$ expresses the finite area of the infinite domain bounded by the graph of the function $f(x)$ from above, by the $x$-axis from below and by the straight line $x = a$ from the left (Fig. 116).

We can give a similar interpretation for integrals (3) and (4).

Let us consider some examples of calculating improper integrals of the first kind.

**Example 1.** $\int_0^{+\infty} \frac{dx}{1+x^2} = \lim_{R\to+\infty} \int_0^R \frac{dx}{1+x^2} = \lim_{R\to+\infty} \arctan x \Big|_0^R =$ $\lim_{R\to+\infty} \arctan R = \frac{\pi}{2}$, i.e. this integral converges.

**Example 2.** $\int_0^{+\infty} \cos x\,dx = \lim_{R\to+\infty} \int_0^R \cos x\,dx = \lim_{R\to+\infty} \sin x \Big|_0^R =$ $\lim_{R\to+\infty} \sin R$, but there is no limit for the function $\sin R$ as $R \to +\infty$ and, consequently, the integral diverges.

**Example 3.** $\int_{-\infty}^{+\infty} e^x\,dx = \int_{-\infty}^{0} e^x\,dx + \int_0^{+\infty} e^x\,dx$; the integral diverges since

$$\int_0^{+\infty} e^x\,dx = \lim_{R\to+\infty} \int_0^R e^x\,dx = \lim_{R\to+\infty} e^x \Big|_0^R = \lim_{R\to+\infty} (e^R - 1) = \infty.$$

**Example 4.** $\int\limits_1^{+\infty} \frac{dx}{x^\alpha}$, $\alpha$ is a number.

(1) If $\alpha \neq 1$, then for any $R > 0$ we have

$$\lim_{R\to+\infty} \int\limits_1^R \frac{dx}{x^\alpha} = \lim_{R\to+\infty} \frac{x^{1-\alpha}}{1-\alpha}\Big|_1^R = \lim_{R\to+\infty} \frac{R^{1-\alpha}-1}{1-\alpha} = \begin{cases} \frac{-1}{1-\alpha} & \text{for } \alpha > 1, \\ \infty & \text{for } \alpha < 1. \end{cases}$$

(2) If $\alpha = 1$, then for any $R > 0$ we have

$$\lim_{R\to+\infty} \int\limits_1^R \frac{dx}{x} = \lim_{R\to+\infty} \ln x\Big|_1^R = \lim_{R\to+\infty} \ln R = \infty.$$

Thus this integral converges for $\alpha > 1$ and diverges for $\alpha \leqslant 1$.

Note that in these examples the calculation of the improper integral was based on its definition.

**2. Improper integrals of unbounded functions.**

**Definition.** *Let the function $f(x)$ be defined on the half-open interval $[a, b)$. The point $x = b$ is said to be singular if the function $f(x)$ is unbounded in any neighbourhood of this point but bounded on any closed interval $[a, b-\varepsilon]$ belonging to $[a, b)$* (**Fig. 117**). *Assume that the function $f(x)$ is integrable on any interval $[a, b-\varepsilon]$, i.e. there is a definite integral $\int\limits_a^{b-\varepsilon} f(x)\,dx$ for any $\varepsilon > 0$ such that $b - \varepsilon > a$. If there is a finite limit*

$$\lim_{\varepsilon\to 0+} \int\limits_a^{b-\varepsilon} f(x)\,dx, \tag{5}$$

*then it is called an improper integral of the second kind and is designated as*

$$\int\limits_a^b f(x)\,dx. \tag{6}$$

In this case we say that integral (6) exists or converges. Now if limit (5) does not exist or is infinite, we say that integral (6) does not exist or diverges.

Similarly, if $x = a$ is a singular point, then the improper integral is defined as follows:

$$\int\limits_a^b f(x)\,dx = \lim_{\varepsilon\to 0+} \int\limits_{a+\varepsilon}^b f(x)\,dx.$$

If the function $f(x)$ is unbounded in the neighbourhood of some interior point $c \in [a, b]$, then, under the condition of the existence

of both integrals from the right, we set, by definition,

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

Finally, if $a$ and $b$ are singular points, then, provided that both integrals from the right exist, the improper integral is defined as the sum

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx,$$

where $c$ is any point from $(a, b)$.

**Example 5.** $\int_0^1 \frac{dx}{x^\alpha}$, $\alpha > 0$ is a number.

(1) If $\alpha \neq 1$, then

$$\lim_{\varepsilon\to 0+} \int_\varepsilon^1 \frac{dx}{x^\alpha} = \lim_{\varepsilon\to 0+} \frac{x^{1-\alpha}}{1-\alpha}\Big|_\varepsilon^1 = \lim_{\varepsilon\to 0+} \frac{1-\varepsilon^{1-\alpha}}{1-\alpha} = \begin{cases} \infty & \text{for } \alpha > 1, \\ \frac{1}{1-\alpha} & \text{for } 0 < \alpha < 1. \end{cases}$$

(2) If $\alpha = 1$, then

$$\lim_{\varepsilon\to 0+} \int_\varepsilon^1 \frac{dx}{x} = \lim_{\varepsilon\to 0+} \ln x\Big|_\varepsilon^1 = \lim_{\varepsilon\to 0+} (-\ln \varepsilon) = \infty.$$

Thus this integral converges for $0 < \alpha < 1$ and diverges for $\alpha \geqslant 1$.

**3. Test for convergence of improper integrals.** Let us discuss the problem of convergence of improper integrals of the form $\int_a^{+\infty} f(x)\,dx$.

**Theorem 8.9 (test for comparison of improper integrals).** *If the functions $f(x)$ and $g(x)$ are continuous on the half-line $[a, +\infty)$ and satisfy the condition $0 \leqslant f(x) \leqslant g(x)$ on it, then the convergence of the integral*

$$\int_a^{+\infty} g(x)\,dx \tag{7}$$

*implies the convergence of the integral*

$$\int_a^{+\infty} f(x)\,dx, \tag{8}$$

*and the divergence of integral* (8) *implies the divergence of integral* (7).

*Proof*. We introduce the designations

$$F(R)=\int_a^R f(x)\,dx,\quad G(R)=\int_a^R g(x)\,dx.$$

Since $0\leqslant f(x)\leqslant g(x)$ for $x\geqslant a$, it follows, by virtue of estimations 1° and 2° (see 1.8.5), that the inequalities

$$0\leqslant F(R)\leqslant G(R) \tag{9}$$

hold true for $R\geqslant a$ and, in addition, the function $F(R)$ (and also $G(R)$) is nondecreasing on the half-line $[a, +\infty)$. Indeed, if $a\leqslant R_1\leqslant R_2$, then $\int_{R_1}^{R_2} f(x)\,dx\geqslant 0$ and, consequently,

$$F(R_2)=\int_a^{R_2} f(x)\,dx=\int_a^{R_1} f(x)\,dx+\int_{R_1}^{R_2} f(x)\,dx\geqslant\int_a^{R_1} f(x)\,dx=F(R_1).$$

Assume that integral (7) converges, i.e. the function $G(R)$ has a finite limit as $R\to+\infty$. From this, since the function $G(R)$ is nondecreasing, it follows that the function $G(R)$ is bounded on $[a, +\infty)$. But then, in accordance with relation (9), the function $F(R)$ is also bounded on $[a, +\infty)$ and, consequently, it has the least upper bound on $[a, +\infty)$. Let $\sup\limits_{[a, +\infty)} F(R)=I$. By the definition of the least upper bound, for any $\varepsilon>0$ there is $R_\varepsilon\geqslant a$ such that $0\leqslant I-F(R_\varepsilon)<\varepsilon$. Since the function $F(R)$ does not decrease on $[a, +\infty)$, the inequality $F(R)\geqslant F(R_\varepsilon)$ is satisfied for any $R>R_\varepsilon$, and, hence, $0\leqslant I-F(R)<\varepsilon$ for $R\geqslant R_\varepsilon$.

Thus

$$|F(R)-I|<\varepsilon\quad\text{for}\quad R\geqslant R_\varepsilon.$$

This means that $\lim\limits_{R\to+\infty} F(R)=I$, i.e. integral (8) converges.

Assume now that integral (8) diverges. Then, if we assume that integral (7) converges, it follows, according to what we have proved above, that integral (8) converges, and this contradicts the hypothesis. Consequently, integral (7) also diverges. ■

**Remark.** We can formulate a similar test for comparison of improper integrals of the second kind as follows: *if the functions $f(x)$ and $g(x)$ are continuous on the half-open interval $(a, b]$ and the conditions $0\leqslant f(x)\leqslant g(x)$ are fulfilled for all points $x$ from an interval $(a, a+\varepsilon)$, then the convergence of the integral $\int_a^b g(x)\,dx$ implies*

*the convergence of the integral* $\int_a^b f(x)\,dx$, *and the divergence of the integral* $\int_a^b f(x)\,dx$ *implies the divergence of the integral* $\int_a^b g(x)\,dx$.

**Example 6.** Investigate the convergence of the integral $\int_1^{+\infty} \frac{dx}{x^2(1+e^{-x})}$.

*Solution.* We compare the integrand function $1/[x^2(1+e^{-x})]$ with the function $1/x^2$ on the half-line $1 \leqslant x < +\infty$. It is evident that

$$\frac{1}{x^2(1+e^{-x})} < \frac{1}{x^2}.$$

But the integral $\int_1^{+\infty} \frac{dx}{x^2}$ converges since $\alpha = 2 > 1$ (see Example 4) and, consequently, according to the test for comparison, the given integral converges too.

**Example 7.** Investigate the convergence of the integral $\int_1^{+\infty} \frac{\sqrt{x}}{1+x}\,dx$.

*Solution.* Comparing the integrand function $\sqrt{x}/(1+x)$ with the function $1/(2\sqrt{x})$ on the half-line $1 \leqslant x < +\infty$, we have

$$\frac{1}{2\sqrt{x}} = \frac{\sqrt{x}}{x+x} \leqslant \frac{\sqrt{x}}{1+x}.$$

But the integral $\int_1^{+\infty} \frac{dx}{2\sqrt{x}}$ diverges since $\alpha = 1/2 < 1$ (see Example 4) and, consequently, according to the test for comparison, the given integral diverges too.

**4. An example showing how the improper integral can be used.** We shall calculate the escape velocity of a body, i.e. the initial velocity at which it can leave the field of attraction of the Earth for space.

Some time earlier (see 1.8.10, Item 6, Example 11) we have used the definite integral to calculate the work needed to launch a body of mass $m$ from the surface of the Earth to the height $h$:

$$A = \int_R^{R+h} F(x)\,dx = \frac{PRh}{R+h}.$$

The escape of a body to space means that it is launched to an infinite height ($h = \infty$). We shall calculate the work needed to do that:

$$\lim_{h\to\infty} A = \int_R^\infty F(x)\,dx = \lim_{h\to\infty} \frac{PRh}{R+h} = \lim_{h\to\infty} \frac{PR}{1+R/h} = PR = mgR,$$

where $m$ is the mass of the body and $g$ is the acceleration of free fall at the surface of the Earth (the friction and the attraction of other planets are neglected). This work is done at the expense of the variation of the kinetic energy of the body. Therefore at the initial moment the kinetic energy of the body must not be smaller than the work, i.e. the initial velocity $v$ of the body must be such that

$$\frac{mv^2}{2} \geqslant mgR, \text{ or } v \geqslant \sqrt{2gR} = \sqrt{2\cdot 10\cdot 6\,400\,000} \text{ m/s}$$

$$= 1.4\cdot 8000 \text{ m/s} = 11.2 \text{ km/s}.$$

If the initial velocity of the body is 11.2 km/s, then the trajectory of its motion is a parabola. At the initial velocity greater than 11.2 km/s the trajectory is a hyperbola and at the initial velocity smaller than 11.2 km/s the body will move along an elliptic path, and will either fall on the Earth or become an Earth satellite.

### 1.8.12. Approximation of Definite Integrals

When solving physical and engineering problems, we have to find definite integrals of functions whose antiderivatives cannot be expressed in terms of elementary functions, and so a necessity arose of deriving approximate formulas for calculation of definite integrals. We shall get acquainted with two of them: the trapezoid formula and the parabolic formula (Simpson's rule).

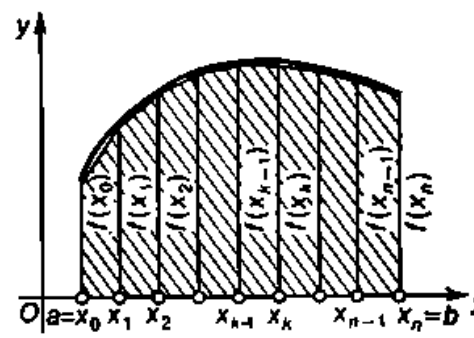


Fig. 118

**1. The trapezoid formula.** Assume that we have to calculate the integral $\int_a^b f(x)\,dx$, where $f(x)$ is a continuous function. To make the reasoning simpler, we restrict our discussion to the case when $f(x) \geqslant 0$. We divide the closed interval $[a, b]$ into $n$ equal subintervals by the points $a = x_0 < x_1 < x_2 < \ldots < x_{k-1} < x_k < \ldots x_{n-1} < x_n = b$ and, using the straight lines $x = x_k$, construct $n$ rectilinear trapezoids (they are hatched in Fig. 118). The sum of the areas of the trapezoids is approximately equal to the area of the

curvilinear trapezoid, i.e.

$$\int_a^b f(x)\,dx \cong \frac{f(x_0)+f(x_1)}{2}(x_1-x_0)+\frac{f(x_1)+f(x_2)}{2}(x_2-x_1)+$$

$$\ldots+\frac{f(x_{n-1})+f(x_n)}{2}(x_n-x_{n-1})=\frac{b-a}{2n}\left[f(a)+f(b)+2\sum_{k=1}^{n-1}f(x_k)\right],$$

where $f(x_{k-1})$ and $f(x_k)$ are the bases of the trapezoids respectively and $x_k - x_{k-1} = (b-a)/n$ are their altitudes.

We have thus obtained an approximate formula

$$\int_a^b f(x)\,dx \cong \frac{b-a}{2n}\left[f(a)+f(b)+2\sum_{k=1}^{n-1}f(x_k)\right],$$

which is known as the *trapezoid formula*. It is the more exact, the larger $n$ is.

Let us consider an integral $\int_0^1 x^2\,dx$ for example. It is easy to find its exact value:

$$\int_0^1 x^2\,dx=\frac{x^3}{3}\Big|_0^1=\frac{1}{3}=0.3333\ldots\ .$$

We shall now calculate its approximate value using the trapezoid formula. Let $n = 5$. Then we have $a = x_0 = 0$, $x_1 = 0.2$, $x_2 = 0.4$, $x_3 = 0.6$, $x_4 = 0.8$, $x_5 = 1 = b$ and, correspondingly, $f(x_0) = 0$, $f(x_1) = 0.04$, $f(x_2) = 0.16$, $f(x_3) = 0.36$, $f(x_4) = 0.64$, $f(x_5) = 1$ and, consequently,

$$\int_0^1 x^2\,dx \cong \frac{1}{10}[1+2(0.04+0.16+0.36+0.64)]=\frac{3.4}{10}=0.34.$$

The exact value of the integral is 0.3333... and therefore the absolute error is smaller than 0.007. This accuracy is sufficient for many engineering problems.

If we increase the number $n$, the accuracy will be higher. Thus, for instance, for $n = 10$

$$\int_0^1 x^2\,dx \cong \frac{1}{20}[1+2\cdot 2.85]=\frac{6.7}{20}=0.335,$$

i.e. the absolute error is smaller than 0.002.

In more complete courses in higher mathematics it is proved that if the function $f(x)$ has a continuous second derivative on the closed

interval $[a, b]$, then the absolute value of the error of the trapezoid formula is not larger than

$$k\frac{(b-a)^3}{12n^2},$$

where $k$ is the greatest value of $|f''(x)|$ on the interval $[a, b]$.

It should be pointed out that with an increase in $n$, not only the accuracy of calculation of the definite integral becomes higher, but the volume of calculations also increases. But in this case computers are used.

We shall use the trapezoid formula to calculate the integral $\int_0^1 \frac{dx}{1+x}$ for $n = 10$. We divide the interval $[0, 1]$ into ten equal parts by the points $x_0 = 0$, $x_1 = 0.1, \ldots, x_9 = 0.9$, $x_{10} = 1$ and find the approximate values of the function $f(x) = 1/(1+x)$ at these points: $f(0) = 1.0000$, $f(0.1) = 0.9091$, $f(0.2) = 0.8333$, $f(0.3) = 0.7692$, $f(0.4) = 0.7143$, $f(0.5) = 0.6667$, $f(0.6) = 0.6250$, $f(0.7) = 0.5882$, $f(0.8) = 0.5556$, $f(0.9) = 0.5263$, $f(1) = 0.5000$.

From the trapezoid formula we obtain

$$\int_0^1 \frac{dx}{1+x} \cong \frac{1}{10}\left(\frac{1.0000+0.5000}{2}+0.9091+0.8333\right.$$
$$\left.+0.7692+0.7143+0.6667+0.6250+0.5882+0.5556+0.5263\right)$$
$$=0.69377 \cong 0.6938.$$

We estimate the error of the result obtained. Since $f(x) = 1/(1+x)$, we have $f'(x) = -1/(1+x)^2$, $f''(x) = 2/(1+x)^3$. On the interval $[0, 1]$ we have $|f''(x)| \leqslant 2$. Therefore the error of the result obtained does not exceed

$$\frac{k(b-a)^3}{12n^2}=\frac{2}{12\cdot 10^2}=\frac{1}{600}<0.0017.$$

We use the Newton-Leibniz formula to find the exact value of this integral:

$$\int_0^1 \frac{dx}{1+x}=\ln(1+x)\Big|_0^1=\ln 2 \cong 0.69315.$$

The absolute error of the result obtained by the trapezoid formula is smaller than 0.0007. This corresponds to the estimate of the error given above.

The idea that was used to construct the trapezoid formula can also be used to obtain more exact approximate formulas for calculation of the definite integral.

**2. Simpson's rule.** We shall first prove two lemmas.

**Lemma 8.1.** *A single curve of the form*

$$y = Ax^2 + Bx + C \tag{1}$$

*can be drawn through any three points* $M_1(x_1, y_1)$, $M_2(x_2, y_2)$, $M_3(x_3, y_3)$ *with different abscissas.*

*Proof.* Substituting the coordinates of the points $M_1$, $M_2$, $M_3$ into the parabolic formula (1), we get a system of three first-degree equations with three unknowns $A$, $B$, $C$:

$$\begin{cases} Ax_1^2 + Bx_1 + C = y_1, \\ Ax_2^2 + Bx_2 + C = y_2, \\ Ax_3^2 + Bx_3 + C = y_3. \end{cases}$$

Since the numbers $x_1$, $x_2$, $x_3$ are distinct, the determinant of this system (see 1.10.3) is nonzero:

$$\Delta = \begin{vmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ x_3^2 & x_3 & 1 \end{vmatrix} = (x_2 - x_1)(x_3 - x_2)(x_1 - x_2) \neq 0.$$

Consequently, this system has a unique solution, i.e. the coefficients $A$, $B$, $C$ are uniquely defined. ■

Note that if $A \neq 0$, then curve (1) is a parabola, and if $A = 0$, then it is a straight line.

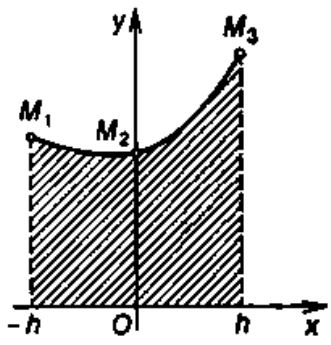



Fig. 119

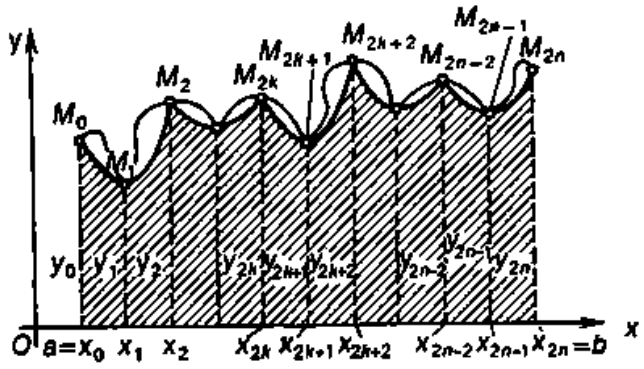



Fig. 120

**Lemma 8.2.** *The area s of the curvilinear trapezoid bounded by the curve* $y = Ax^2 + Bx + C$ *which passes through the points* $M_1(-h, y_1)$, $M_2(0, y_2)$, $M_3(h, y_3)$ (Fig. 119) *is expressed by the formula*

$$s = \frac{h}{3}(y_1 + 4y_2 + y_3). \tag{2}$$

*Proof.* Substituting the coordinates of the points $M_1$, $M_2$, $M_3$ into the equation $y = Ax^2 + Bx + C$, we get $y_1 = Ah^2 - Bh + C$,

$y_2 = C$, $y_3 = Ah^2 + Bh + C$, whence it follows that

$$2Ah^2 + 2C = y_1 + y_3, \quad C = y_2. \tag{3}$$

Taking relations (3) into account, we obtain

$$s = \int_{-h}^{h} (Ax^2 + Bx + C)\,dx = \int_{-h}^{h} (Ax^2 + C)\,dx + B\int_{-h}^{h} x\,dx$$

$$= 2\int_{0}^{h} (Ax^2 + C)\,dx = \frac{h}{3}(2Ah^2 + 6C) = \frac{h}{3}(y_1 + 4y_2 + y_3). \quad \blacksquare$$

Let us again consider the curvilinear trapezoid bounded by the arbitrary curve $y = f(x)$. We divide the interval $[a, b]$ into $2n$ equal subintervals by the points $a = x_0 < x_1 < x_2 < \ldots < x_{2k} < x_{2k+1} < x_{2k+2} < \ldots < x_{2n-2} < x_{2n-1} < x_{2n} = b$ and, using the straight lines $x = x_k$, we divide the curve $y = f(x)$ into $2n$ corresponding parts by the points $M_0, M_1, M_2, \ldots, M_{2k}, M_{2k+1}, M_{2k+2}, \ldots, M_{2n-2}, M_{2n-1}, M_{2n}$ (Fig. 120). Through every triple of points

$$M_0M_1M_2, \ldots, M_{2k}M_{2k+1}M_{2k+2}, \ldots, M_{2n-2}M_{2n-1}M_{2n}$$

we draw a curve of the form $y = Ax^2 + Bx + C$ (see Lemma 8.1). As a result, we get $n$ curvilinear trapezoids bounded by parabolas or straight lines from above (these trapezoids are hatched in Fig. 120). Since the area of the partial curvilinear trapezoid which corresponds to the subinterval $[x_{2k}, x_{2k+2}]$ is approximately equal to the area of the corresponding "parabolic" trapezoid, we have from formula (2) (in this case $h = (b - a)/(2n)$)

$$\int_{x_{2k}}^{x_{2k+2}} f(x)\,dx \cong \frac{b-a}{6n}(y_{2k} + 4y_{2k+1} + y_{2k+2}),$$

where $y_k = f(x_k)$, $k = 0, 1, 2, \ldots, 2n$. Adding these approximate equalities term by term, we get an approximate formula

$$\int_{a}^{b} f(x)\,dx \cong \frac{b-a}{6n}\sum_{k=0}^{n-1}(y_{2k} + 4y_{2k+1} + y_{2k+2})$$

or, in full,

$$\int_{a}^{b} f(x)\,dx \cong \frac{b-a}{6n}[y_0 + y_{2n} + 2(y_2 + y_4 + \ldots + y_{2n-2}) + 4(y_1 + y_3 + \ldots + y_{2n-1})].$$

This formula is known as the *parabolic formula*, or *Simpson's* rule*.

* Simpson, Thomas (1710-1761), an English mathematician.

In this formula the value of the function $f(x)$ at odd division points $x_1, x_3, \ldots, x_{2n-1}$ has a coefficient 4, at even points $x_2, x_4, \ldots, x_{2n-2}$ it has a coefficient 2, and at two boundary points $a = x_0$ and $x_{2n} = b$ it has a coefficient 1.

The geometric meaning of Simpson's rule is obvious: the area of the curvilinear trapezoid under the graph of the function $f(x)$ on the interval $[a, b]$ can be approximated by the sum of the areas of the figures lying under the parabolas (straight lines).

In complete courses in higher mathematics it is proved that if the function $f(x)$ has a forth-order continuous derivative on $[a, b]$, then the absolute value of the error of Simpson's formula is not greater than

$$M\frac{(b-a)^5}{2880n^4},$$

where $M$ is the greatest value of $|f^4(x)|$ on the interval $[a, b]$. As was pointed out above, the error of the trapezoid formula is estimated as

$$k\frac{(b-a)^3}{12n^2} \quad \left(k = \max_{[a,\,b]} |f''(x)|\right).$$

Since $n^4$ increases faster than $n^2$, the error of Simpson's formula decreases, with an increase in $n$, considerably faster than the error of the trapezoid formula. This explains why Simpson's formula makes it possible to obtain a higher accuracy than the trapezoid formula.

To compare the accuracy of the approximate formulas, we shall again calculate the integral $\int_0^1 \frac{dx}{1+x}$, but now we shall do this by Simpson's formula for $n = 4$. We divide the interval $[0, 1]$ into four equal parts by the points $x_0 = 0$, $x_1 = 1/4$, $x_2 = 1/2$, $x_3 = 3/4$, $x_4 = 1$ and find the approximate values of the function $f(x) = 1/(1 + x)$ at these points: $y_0 = 1.0000$, $y_1 = 0.8000$, $y_2 = 0.6667$, $y_3 = 0.5714$, $y_4 = 0.5000$.

From Simpson's formula we get

$$\int_0^1 \frac{dx}{1+x} \cong \frac{b-a}{6n}[y_0 + y_4 + 2y_2 + 4(y_1 + y_3)]$$

$$= \frac{1-0}{12}[1.0000 + 0.5000 + 2\cdot 0.6667 + 4(0.8000 + 0.5714)] \cong 0.69325.$$

We estimate the error of the result obtained. For the integrand function $f(x) = 1/(1 + x)$ we have $f^{(4)}(x) = 24/(1 + x)^5$, whence it follows that $|f^{(4)}(x)| \leqslant 24$ on the interval $[0, 1]$. Consequently, we can take $M = 24$ and the error of the result does not exceed

$24/(2880 \cdot 4^4) < 0.0004$. Comparing the approximate value with the exact one, we infer that the absolute error of the result obtained by Simpson's formula is smaller than 0.00011. This corresponds to the estimate of the error given above and, in addition, testifies to the fact that Simpson's formula is much more exact than the trapezoid formula. That is why Simpson's formula is more often used to calculate definite integrals than the trapezoid formula.

As was pointed out above, the approximate formulas are used to calculate definite integrals when the antiderivative of the integrand function cannot be expressed in terms of elementary functions.

Let us use Simpson's formula to calculate, for instance, the integral $\int_0^1 e^{-x^2}\,dx$* with an accuracy to within 0.001.

To choose the number $2n$ necessary to obtain the desired degree of accuracy, we seek $f^{(4)}(x)$. Successively differentiating the function $f(x) = e^{-x^2}$, we get

$$f^{(4)}(x) = 4e^{-x^2}(4x^4 - 12x^2 + 3).$$

Since $e^{-x^2} \leqslant 1$ and $|4x^4 - 12x^2 + 3| \leqslant 5$ on the interval $[0, 1]$, we have $|f^{(4)}(x)| \leqslant 20$. Consequently, we can take $M = 20$. Using the formula for the estimate of the error, we have $20/2880n^4 < 1/1000$, whence $n^4 > 1000/144$. For this inequality to hold true, it is sufficient to take $n = 2$, i.e. $2n = 4$.

Now we divide the interval $[0, 1]$ into four equal parts by the points $x_0 = 0$, $x_1 = 1/4$, $x_2 = 1/2$, $x_3 = 3/4$, $x_4 = 1$ and find the approximate values of the function $f(x) = e^{-x^2}$ at these points: $y_0 = 1.0000$, $y_1 = 0.9394$, $y_2 = 0.7788$, $y_3 = 0.5698$, $y_4 = 0.3679$. Using Simpson's formula, we obtain

$$\int_0^1 e^{-x^2}\,dx \cong \frac{1}{12}[1.0000 + 0.3679 + 2 \cdot 0.7788$$

$$+ 4(0.9394 + 0.5698)] \cong 0.7469.$$

Thus $\int_0^1 e^{-x^2}\,dx \cong 0.747$ with an accuracy to within 0.001. Having divided the interval $[0, 1]$ only into four equal parts and replaced the integral in question by the sum appearing on the right-hand side of Simpson's formula, we have calculated the given integral with the required accuracy.

---

* This integral cannot be expressed in terms of elementary functions but is very significant in statistical physics, the theory of heat conduction and diffusion.

It should be pointed out in conclusion that each of the methods of approximation of integrals we have discussed contains a strict algorithm for finding them, and this makes it possible to use these methods for computer calculations. Thus these methods are an efficient means for calculation of integrals. For integrals, which cannot be expressed in terms of elementary functions, we can compile tables of their values using computers and the simplest methods of approximation.

# Part Two

# ANALYSIS OF FUNCTIONS OF SEVERAL VARIABLES

Before considering functions of two and more variables, we shall get acquainted with some notions of solid analytic geometry which will be needed later on for geometric interpretation of functions of several variables.

In this part we shall also consider such concepts of higher algebra as the matrix and the determinant which are not only important for the study of analysis but also have a variety of applications in other divisions of mathematics.

## Chapter 9

## SOLID ANALYTIC GEOMETRY

### 2.9.1. Rectangular Coordinate System in Space

The rectangular system of coordinates $Oxyz$ in space is defined by specifying a scale unit for measuring lengths and three mutually perpendicular axes $Ox$, $Oy$, $Oz$, which meet at a point $O$. The point $O$ is the *origin of coordinates*, or simply *origin*, $Ox$ is the *abscissa axis*, $Oy$ is the *axis of ordinates*, and $Oz$ is the *axis of applicates*.

Let $M$ be an arbitrary point of space (Fig. 121). We draw three planes perpendicular to the axes $Ox$, $Oy$, $Oz$ through the point $M$. The points of intersection of the planes and the axes will be designated as $M_x$, $M_y$, $M_z$ respectively. The *rectangular coordinates of the point* $M$ are the numbers

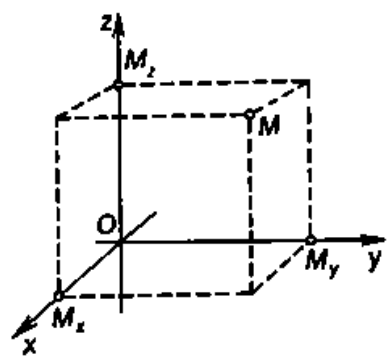



Fig. 121

$$x = OM_x, \quad y = OM_y, \quad z = OM_z,$$

i.e. the values of the directed segments $\overrightarrow{OM_x}$, $\overrightarrow{OM_y}$, $\overrightarrow{OM_z}$, where $x$ is the *abscissa*, $y$ is the *ordinate* and $z$ is the *applicate* of the point $M$.

Thus, in the chosen system of coordinates, every point $M$ of space is associated with a single ordered triple of numbers $(x, y, z)$* which are its rectangular coordinates and, conversely, every ordered triple of numbers $(x, y, z)$ is associated with a point $M$ in space and that point is unique.

Thus a rectangular coordinate system in space establishes a one-to-one correspondence between the set of all points of space and the set of ordered triples of numbers.

The planes $Oxy$, $Oyz$, $Oxz$ are *coordinate planes*. They divide the whole space into eight parts called *octants*.

### 2.9.2. The Vector

**1. Scalar and vector quantities.** Many physical quantities, such, for instance, as volume, mass, density, the temperature of a body, can be completely defined by a number. Quantities of this kind are *scalar quantities*, and therefore numbers are sometimes called scalars. But there are quantities which are defined not only by a number but also by a direction. For instance, for a moving body not only the velocity must be specified but also the direction of its motion. When the action of a force is studied, not only its value must be indicated but also the direction of its application. Quantities of this kind are *vector quantities*. To describe them, the concept of a vector was introduced which proved to be useful for mathematics.

**2. Definition of the vector.** Any ordered pair of points $A$ and $B$ of space defines a *directed segment*, i.e. a line segment with a direction specified on it. If the point $A$ is the first, then it is the *origin* of the directed segment, and the point $B$ is its *terminal point*, or *terminus*. The direction of the segment is that from the origin to the terminal point.

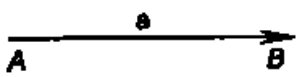


Fig. 122

**Definition 1.** *A directed segment is a vector.*

We denote a vector by the symbol $\overrightarrow{AB}$, the first letter representing the origin and the second, the terminus of the vector. A vector can also be denoted by one letter written in bold-face type, **a**, for instance (Fig. 122).

A vector whose origin and terminus coincide is a *zero*, or *null*, *vector* and is designated as **0** or simply 0.

The distance between the origin and the terminus of a vector is its *length* and is designated as $|\overrightarrow{AB}|$ or $|\mathbf{a}|$.

* The triple of numbers $x$, $y$, $z$ is said to be ordered if it is indicated which of these numbers is the first, which is the second and which is the third. In the notation $(x, y, z)$, $x$ is the first number, $y$ is the second and $z$ is the third.

The vectors **a** and **b** are *collinear* if they lie on the same straight line or on parallel straight lines. Collinear vectors may be of the same direction or of opposite directions.

A zero vector is considered to be of the same direction as any vector; its length is zero, i.e. $|\mathbf{0}| = 0$.

We can formulate now an important concept of the equality of two vectors.

**Definition 2.** *The vectors* **a** *and* **b** *are equal* (**a** = **b**) *if they are collinear, of the same direction and equal in length.*

Figure 123 shows unequal (on the left) and equal (on the right) vectors **a** and **b**. From the definition of the equality of vectors it

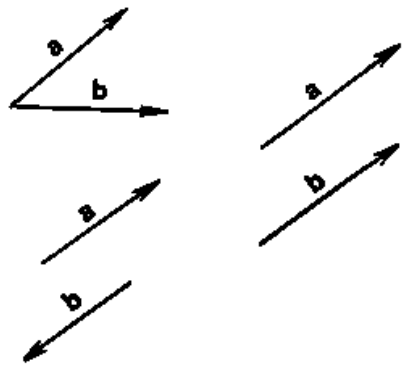



Fig. 123

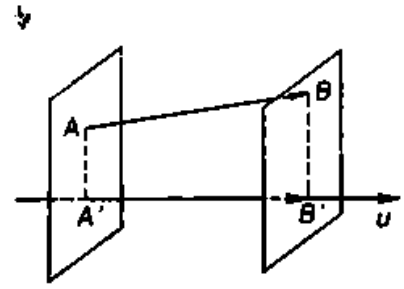



Fig. 124

follows that if a vector is displaced parallel to itself, a vector results which is equal to the given vector. That is why in analytic geometry vectors are said to be *free*.

**3. The projection of a vector on an axis.** Assume that an axis $u$ and a vector $\overrightarrow{AB}$ are given in space. Through the points $A$ and $B$ we draw planes perpendicular to the $u$-axis and designate the points of intersection of the planes and the axis as $A'$ and $B'$ (Fig. 124).

The projection of the vector $\overrightarrow{AB}$ on the $u$-axis is the value $A'B'$ of the directed segment $\overrightarrow{A'B'}$ on the $u$-axis. Recall that

$$\left.\begin{array}{ll} A'B' = |\overrightarrow{A'B'}| & \text{if the direction of } \overrightarrow{A'B'} \text{ coincides with that of the } u\text{-axis,} \\ A'B' = -|\overrightarrow{A'B'}| & \text{if the direction of } \overrightarrow{A'B'} \text{ is opposite to that of the } u\text{-axis.} \end{array}\right\} \quad (1)$$

The projection of the vector $\overrightarrow{AB}$ on the $u$-axis is designated as $\text{proj}_u\, \overrightarrow{AB}$.

Consider the following theorem.

**Theorem 9.1.** *The projection of the vector $\overrightarrow{AB}$ on the u-axis is equal to the length of the vector $\overrightarrow{AB}$ multiplied by the cosine of the angle between the vector $\overrightarrow{AB}$ and the u-axis, i.e.*

$$\operatorname{proj}_u \overrightarrow{AB} = |\overrightarrow{AB}| \cos \varphi, \tag{2}$$

*where $\varphi$ is the angle between the vector $\overrightarrow{AB}$ and the u-axis* (Fig. 125).

*Proof.* If $\varphi \leqslant \pi/2$ (Fig. 125*a*), then, by virtue of (1),

$$\operatorname{proj}_u \overrightarrow{AB} = |\overrightarrow{A'B'}| = |\overrightarrow{AB}| \cos \varphi.$$

Now if $\varphi > \pi/2$ (Fig. 125*b*), then, by virtue of (1),

$$\operatorname{proj}_u \overrightarrow{AB} = -|\overrightarrow{A'B'}| = -|\overrightarrow{AB}| \cos(\pi - \varphi) = |\overrightarrow{AB}| \cos \varphi.$$

Thus, relation (2) holds true for any angle $\varphi$. ■

**Remark 1.** Assume that $\overrightarrow{A_1B_1} = \overrightarrow{A_2B_2}$ and an axis $u$ is given. Applying relation (2) to each of these vectors, we get

$$\operatorname{proj}_u \overrightarrow{A_1B_1} = \operatorname{proj}_u \overrightarrow{A_2B_2},$$

i.e. equal vectors have equal projections on the same axis.

**4. Projections of vectors on the coordinate axes.** Assume that a rectangular coordinate system $Oxyz$ and an arbitrary vector $\overrightarrow{AB}$

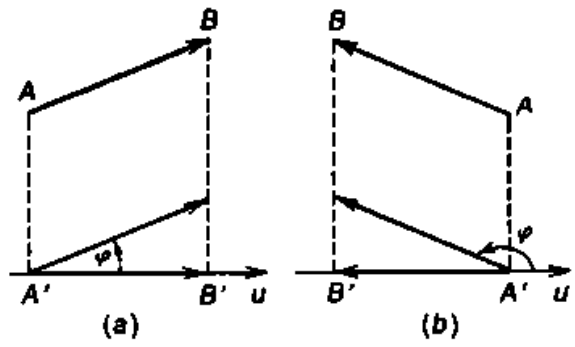



Fig. 125

are given in space. Assume furthermore that $X = \operatorname{proj}_x \overrightarrow{AB}$, $Y = \operatorname{proj}_y \overrightarrow{AB}$, $Z = \operatorname{proj}_z \overrightarrow{AB}$. The projections $X$, $Y$, $Z$ of the vector $\overrightarrow{AB}$ on the coordinate axes are its *coordinates*. In this case we write

$$\overrightarrow{AB} = \{X, Y, Z\}.$$

**Theorem 9.2.** *Whatever the two points $A(x_1, y_1, z_1)$ and $B(x_2, y_2, z_2)$, the coordinates of the vector $\overrightarrow{AB}$ are defined by the following formulas:*

$$X = x_2 - x_1, \quad Y = y_2 - y_1, \quad Z = z_2 - z_1. \tag{3}$$

*Proof.* Through the points $A$ and $B$ we draw planes perpendicular to the $x$-axis and designate the points of their intersection with the

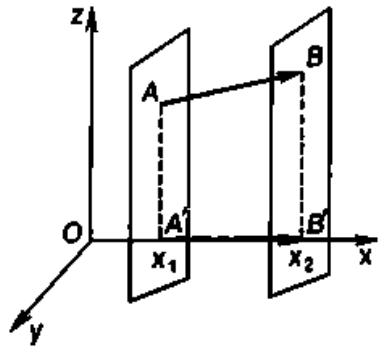



Fig. 126

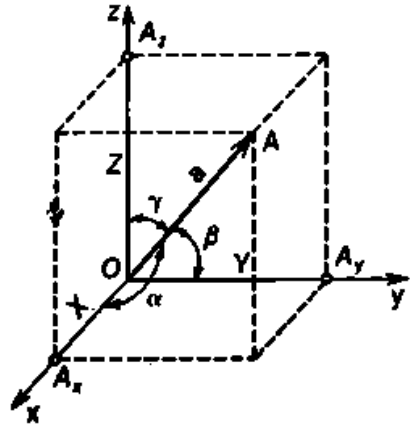



Fig. 127

$x$-axis as $A'$ and $B'$ respectively. The points $A'$ and $B'$ on the $x$-axis have coordinates $x_1$ and $x_2$ (Fig. 126). By definition, $X = \text{proj}_x \overrightarrow{AB} = A'B'$. But $A'B' = x_2 - x_1$ (see 1.1.3), and therefore $X = x_2 - x_1$. We can establish the other formulas (3) by analogy. ■

**Remark 2.** If the vector $\overrightarrow{AB}$ begins at the origin of coordinates, i.e. $x_1 = y_1 = z_1 = 0$, and $x_2 = x$, $y_2 = y$, $z_2 = z$, then the coordinates $X$, $Y$, $Z$ of the vector $\overrightarrow{AB}$ are equal to the coordinates of its terminus, i.e.

$$X = x, \quad Y = y, \quad Z = z.$$

**5. Direction cosines of a vector.** Consider an arbitrary vector $\mathbf{a} = \{X, Y, Z\}$. We assume that $\mathbf{a}$ begins at the origin of coordinates and does not lie in any coordinate plane. Through the point $A$ we draw planes perpendicular to the axes. Together with the coordinate planes they form a right parallelepiped whose diagonal is the segment $OA$ (Fig. 127). It is known from elementary geometry that the square of the length of the diagonal of a right parallelepiped is equal to the sum of the squares of the lengths of its three dimensions. Consequently,

$$|OA|^2 = |OA_x|^2 + |OA_y|^2 + |OA_z|^2.$$

But $|OA| = |\mathbf{a}|$, $|OA_x| = |X|$, $|OA_y| = |Y|$, $|OA_z| = |Z|$. We thus have

$$|\mathbf{a}|^2 = X^2 + Y^2 + Z^2,$$

or

$$|\mathbf{a}| = \sqrt{X^2+Y^2+Z^2}. \tag{4}$$

Formula (4) expresses the length of an arbitrary vector in terms of its coordinates.

We designate the angles between the vector **a** and the coordinate axes as $\alpha$, $\beta$, $\gamma$. From formulas (2) and (4) we have

$$\cos\alpha = \frac{X}{\sqrt{X^2+Y^2+Z^2}},$$
$$\cos\beta = \frac{Y}{\sqrt{X^2+Y^2+Z^2}}, \tag{5}$$
$$\cos\gamma = \frac{Z}{\sqrt{X^2+Y^2+Z^2}},$$

$\cos\alpha$, $\cos\beta$, $\cos\gamma$ are the *direction cosines of the vector* **a**.

Squaring the left-hand and right-hand sides of each of the relations (5) and summing up the results, we obtain

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1, \tag{6}$$

i.e. the sum of the squares of the direction cosines of any vector is equal to unity.

In conclusion, we shall consider a problem. We are given two arbitrary points $M_1(x_1, y_1, z_1)$ and $M_2(x_2, y_2, z_2)$ and have to find the distance $d$ between them. Using Theorem 9.2 and formula (4), we at once get the desired result:

$$\overrightarrow{M_1M_2} = \{x_2-x_1, y_2-y_1, z_2-z_1\},$$

and, since $d$ is the length of the vector $\overrightarrow{M_1M_2}$, it follows that

$$d = |\overrightarrow{M_1M_2}| = \sqrt{(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2}. \tag{7}$$

### 2.9.3. Linear Operations Involving Vectors and Their Principal Properties

The operations of addition and subtraction of vectors, and multiplication of a vector by a scalar are linear operations involving vectors.

**1. Addition of two vectors.** Consider two vectors **a** and **b**. *The sum* **a** + **b** *is the vector from the origin of* **a** *to the terminus of* **b** *provided that the vector* **b** *is applied to the terminus of the vector* **a** (Fig. 128*a*).

**Remark 1.** Subtraction of vectors is the inverse of addition, i.e. *the difference* **b** — **a** *of the vectors* **b** *and* **a** *is the vector which, when added to the vector* **a**, *yields the vector* **b** (Fig. 128*b*).

**Remark 2.** Having defined the sum of two vectors, we can find the sum of any number of vectors. Consider, for instance, three vectors **a**, **b**, **c**. Adding the vectors **a** and **b** together, we get a vector **a** + **b**. Adding the vector **c** to it, we get a vector **a** + **b** + **c**.

**2. The product of a vector by a scalar.** Consider a vector $\mathbf{a} \neq \mathbf{0}$ and a scalar $\lambda \neq 0$. *The product* $\lambda\mathbf{a}$ *is a vector which is collinear with*

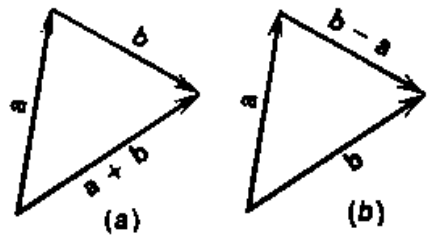



Fig. 128

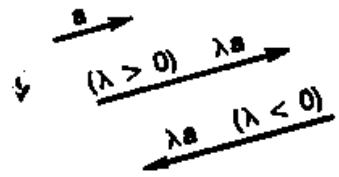



Fig. 129

*the vector* **a**, *has a length equal to* $|\lambda|\,|\mathbf{a}|$, *and the same direction as the vector* **a** *if* $\lambda > 0$ *and the opposite direction if* $\lambda < 0$ (Fig. 129).

We can express the geometric meaning of the operation of multiplication of the vector $\mathbf{a} \neq \mathbf{0}$ by the scalar $\lambda \neq 0$ as follows: if $|\lambda| > 1$, then, when the vector **a** is multiplied by the scalar $\lambda$, the vector **a** is "extended" $\lambda$ times, and if $|\lambda| < 1$, then it is "contracted" $1/\lambda$ times. When $\lambda < 0$, the vector changes its direction to the opposite. Figure 129 shows the case $|\lambda| > 1$.

If $\lambda = 0$ or $\mathbf{a} = \mathbf{0}$, then the product $\lambda\mathbf{a}$ is considered to be equal to a zero vector.

**Remark 3.** Using the definition of the multiplication of a vector by a scalar, it is easy to prove that if the vectors **a** and **b** are collinear and $\mathbf{a} \neq \mathbf{0}$, then there is a number $\lambda$ such that $\mathbf{b} = \lambda\mathbf{a}$ and this number is unique (prove this statement yourself!).

**3. Principal properties of linear operations.**

1°. $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ (*commutativity of addition*).

*Proof.* Applying the vectors **a** and **b** to the same point $O$, we construct a parallelogram on them (Fig. 130). Then

$$\mathbf{a} = \overrightarrow{OA}, \quad \mathbf{b} = \overrightarrow{AC}, \quad \mathbf{a} + \mathbf{b} = \overrightarrow{OC},$$

$$\mathbf{b} = \overrightarrow{OB}, \quad \mathbf{a} = \overrightarrow{BC}, \quad \mathbf{b} + \mathbf{a} = \overrightarrow{OC}.$$

Consequently, $\overrightarrow{OC} = \mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$. ■

2°. $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$ (*associativity of addition*).

*Proof.* We arrange the vectors in question so that **b** is applied to the terminus of **a**, and **c** to the terminus of **b**. We designate the origin

of the vector **a** as $O$, its terminus as $A$, the terminus of **b** as $B$ and the terminus of **c** as $C$ (Fig. 131). Then

$$(\mathbf{a}+\mathbf{b})+\mathbf{c}=(\overrightarrow{OA}+\overrightarrow{AB})+\overrightarrow{BC}=\overrightarrow{OB}+\overrightarrow{BC}=\overrightarrow{OC},$$

$$\mathbf{a}+(\mathbf{b}+\mathbf{c})=\overrightarrow{OA}+(\overrightarrow{AB}+\overrightarrow{BC})=\overrightarrow{OA}+\overrightarrow{AC}=\overrightarrow{OC}.$$

Consequently, $\overrightarrow{OC}=(\mathbf{a}+\mathbf{b})+\mathbf{c}=\mathbf{a}+(\mathbf{b}+\mathbf{c})$. ■

Let us consider three more properties of linear operations, two of which simultaneously refer to the addition of vectors and the multi-

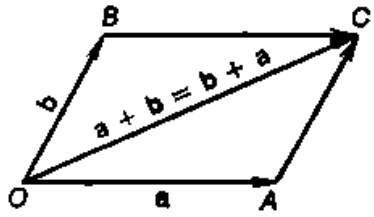



Fig. 130

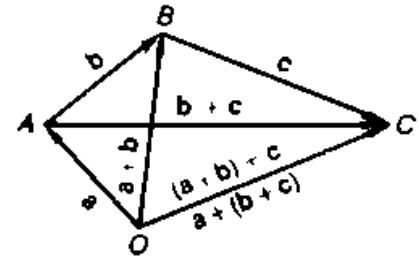



Fig. 131

plication of a vector by a scalar. Let $\lambda$ and $\mu$ be arbitrary numbers and **a** and **b** be any vectors. Then

3°. $\lambda(\mu\mathbf{a})=(\lambda\mu)\mathbf{a}$ (*associativity of multiplication*).

4°. $(\lambda+\mu)\mathbf{a}=\lambda\mathbf{a}+\mu\mathbf{a}$ (*distributivity relative to the sum of numbers*).

5°. $\lambda(\mathbf{a}+\mathbf{b})=\lambda\mathbf{a}+\lambda\mathbf{b}$ (*distributivity relative to the sum of vectors*).

*Proof of Property* 3°. If at least one of the numbers $\lambda$, $\mu$ or the vector **a** is zero, then both sides of relation 3° vanish. If $\lambda\neq 0$, $\mu\neq 0$, $\mathbf{a}\neq\mathbf{0}$, then the vectors $\lambda(\mu\mathbf{a})$ and $(\lambda\mu)\mathbf{a}$ are collinear, have the same direction (their directions either coincide with the direction of the vector **a** if $\lambda$ and $\mu$ are of like sign or are opposite to the direction of the vector **a** if $\lambda$ and $\mu$ are of unlike signs) and have the same length ($|\lambda(\mu\mathbf{a})|=|\lambda||\mu\mathbf{a}|=|\lambda||\mu||\mathbf{a}|$ and $|(\lambda\mu)\mathbf{a}|=|\lambda\mu||\mathbf{a}|=|\lambda||\mu||\mathbf{a}|$) and, consequently, they are equal. ■

*Proof of Property* 4°. Let $\lambda$ and $\mu$ be of like sign and $\mathbf{a}\neq\mathbf{0}$. Then the vectors $(\lambda+\mu)\mathbf{a}$ and $\lambda\mathbf{a}+\mu\mathbf{a}$ are collinear and have the same direction (for $\lambda>0$ and $\mu>0$ their directions coincide with the direction of the vector **a** and for $\lambda<0$ and $\mu<0$ their directions are opposite to that of **a**). Since the vectors $\lambda\mathbf{a}$ and $\mu\mathbf{a}$ have the same direction, the length of the vector $\lambda\mathbf{a}+\mu\mathbf{a}$ is equal to $|\lambda\mathbf{a}+\mu\mathbf{a}|=|\lambda\mathbf{a}|+|\mu\mathbf{a}|=|\lambda||\mathbf{a}|+|\mu||\mathbf{a}|=(|\lambda|+|\mu|)|\mathbf{a}|$, and since $\lambda$ and $\mu$ are of like sign, it follows that $|(\lambda+\mu)\mathbf{a}|=|\lambda+\mu||\mathbf{a}|=(|\lambda|+|\mu|)|\mathbf{a}|$, i.e. the length of the vector $\lambda\mathbf{a}+\mu\mathbf{a}$ is equal to the length of the vector $(\lambda+\mu)\mathbf{a}$. Thus the vectors

$(\lambda + \mu)\,\mathbf{a}$ and $\lambda\mathbf{a} + \mu\mathbf{a}$ are collinear, have the same direction and equal length and, consequently, they are equal.

Assume now that $\lambda$ and $\mu$ are of different signs and, for definiteness, $|\lambda| > |\mu|$. In this case the vectors $(\lambda + \mu)\,\mathbf{a}$ and $\lambda\mathbf{a} + \mu\mathbf{a}$ have the same direction as the vector $\lambda\mathbf{a}$. The length of the vector $\lambda\mathbf{a} + \mu\mathbf{a}$ is equal to $|\lambda\mathbf{a} + \mu\mathbf{a}| = |\lambda\mathbf{a}| - |\mu\mathbf{a}| = |\lambda|\,|\mathbf{a}| - |\mu|\,|\mathbf{a}| = (|\lambda| - |\mu|)\,|\mathbf{a}|$ and $|(\lambda + \mu)\,\mathbf{a}| = |\lambda + \mu|\,|\mathbf{a}| = (|\lambda| - |\mu|)\,|\mathbf{a}|$, i.e. the length of the vector $\lambda\mathbf{a} + \mu\mathbf{a}$ is equal to the length of the vector $(\lambda + \mu)\,\mathbf{a}$. Consequently, in this case as well the vectors $(\lambda + \mu)\,\mathbf{a}$ and $\lambda\mathbf{a} + \mu\mathbf{a}$ are equal. Now if $|\lambda| = |\mu|$ and the numbers $\lambda$ and $\mu$ are of opposite signs, then both sides of the equality being proved are zero. Equality 4° is obvious if at least one of the numbers $\lambda$, $\mu$ or the vector $\mathbf{a}$ is zero. ■

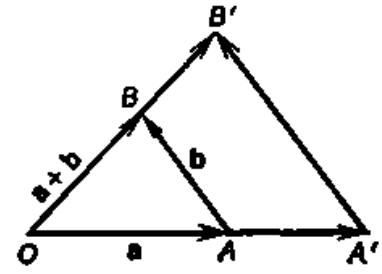


Fig. 132

*Proof of Property* 5°. Let $\mathbf{a}$ and $\mathbf{b}$ be non-collinear vectors and $\lambda > 0$. We construct vectors $\overrightarrow{OB} = \mathbf{a} + \mathbf{b}$, $\overrightarrow{OA'} = \lambda\mathbf{a}$ and $\overrightarrow{OB'} = \lambda(\mathbf{a} + \mathbf{b})$ (Fig. 132). From the similarity of the triangles $OAB$ and $OA'B'$ and from the definition of the operation of multiplication of a vector by a scalar it follows that $\overrightarrow{A'B'} = \lambda\mathbf{b}$, and from the triangle $OA'B'$ we get $\overrightarrow{OA'} + \overrightarrow{A'B'} = \overrightarrow{OB'}$. Thus $\lambda\mathbf{a} + \lambda\mathbf{b} = \lambda(\mathbf{a} + \mathbf{b})$, i.e. the equality being proved holds true.

The case $\lambda < 0$ can be treated by analogy.

If $\mathbf{a}$ and $\mathbf{b}$ are collinear vectors and $\mathbf{a} \neq 0$, then the vector $\mathbf{b}$ can be represented as $\mathbf{b} = \mu\mathbf{a}$, and the equality in question follows from equalities 3° and 4°. Indeed, $\lambda(\mathbf{a} + \mathbf{b}) = \lambda(\mathbf{a} + \mu\mathbf{a}) = \lambda(1 + \mu)\,\mathbf{a} = (\lambda + \lambda\mu)\,\mathbf{a} = \lambda\mathbf{a} + \lambda\mu\mathbf{a} = \lambda\mathbf{a} + \lambda(\mu\mathbf{a}) = \lambda\mathbf{a} + \lambda\mathbf{b}$. The equality being proved is obvious if at least one of the vectors $\mathbf{a}$, $\mathbf{b}$ or the number $\lambda$ is zero. ■

**Remark 4.** The properties of linear operations we have proved are of a fundamental significance since they make it possible to do ordinary algebraic operations on vectors. For instance, by virtue of Properties 4° and 5°, we can multiply a scalar polynomial by a vector polynomial "term by term".

### 2.9.4. Theorems on the Projections of Vectors

**Theorem 9.3.** *The projection of the sum of two vectors on an axis is equal to the sum of their projections on that axis, i.e.*

$$\mathrm{proj}_u\,(\mathbf{a}_1 + \mathbf{a}_2) = \mathrm{proj}_u\,\mathbf{a}_1 + \mathrm{proj}_u\,\mathbf{a}_2.$$

*Proof.* Assume that the points $A$ and $A_1$ are the origin and the terminus of the vector $\mathbf{a}_1$ respectively, the points $A_1$ and $A_2$ are the

origin and the terminus of the vector $\mathbf{a}_2$ (Fig. 133). We designate the projections of the points $A$, $A_1$, $A_2$ on the $u$-axis as $A'$, $A_1'$, $A_2'$ respectively. By definition, $\text{proj}_u\ \mathbf{a}_1 = A'A_1'$, $\text{proj}_u\ \mathbf{a}_2 = A_1'A_2'$, $\text{proj}_u\ (\mathbf{a}_1 + \mathbf{a}_2) = \text{proj}_u\ \overrightarrow{AA_2} = A'A_2'$. According to the fundamental identity (see 1.1.3), $A'A_2' = A'A_1' + A_1'A_2'$. Hence $\text{proj}_u\ (\mathbf{a}_1 + \mathbf{a}_2) = \text{proj}_u\ \mathbf{a}_1 + \text{proj}_u\ \mathbf{a}_2$. ■

This theorem can be generalized to the case of any number of terms.

**Theorem 9.4.** *When the vector* $\mathbf{a}$ *is multiplied by the scalar* $\lambda$, *its projection on an axis is also multiplied by this scalar, i.e.*

$$\text{proj}_u\ \lambda\mathbf{a} = \lambda\ \text{proj}_u\ \mathbf{a}. \tag{1}$$

*Proof.* Assume that $\varphi$ is the angle between the vector $\mathbf{a}$ and the $u$-axis, and $\varphi'$ is the angle between the vector $\lambda\mathbf{a}$ and the $u$-axis

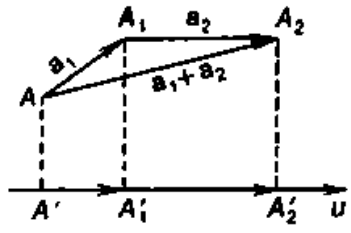



Fig. 133

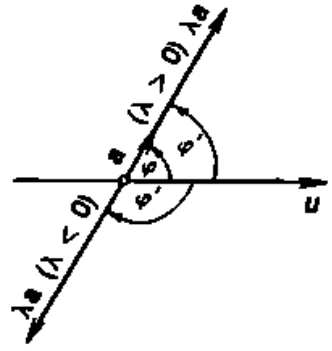



Fig. 134

(Fig. 134). If $\lambda > 0$, then the vectors $\mathbf{a}$ and $\lambda\mathbf{a}$ have the same direction and $\varphi = \varphi'$. Now if $\lambda < 0$, then the vectors $\mathbf{a}$ and $\lambda\mathbf{a}$ are opposite in direction and $\varphi' = \pi - \varphi$. From Theorem 9.1 we have:

for $\lambda > 0$

$$\text{proj}_u\ \lambda\mathbf{a} = |\lambda\mathbf{a}| \cos\varphi' = \lambda\,|\mathbf{a}| \cos\varphi' = \lambda\,|\mathbf{a}| \cos\varphi = \lambda\ \text{proj}_u\ \mathbf{a},$$

for $\lambda < 0$

$$\text{proj}_u\ \lambda\mathbf{a} = |\lambda\mathbf{a}| \cos\varphi' = |\lambda|\,|\mathbf{a}| \cos\varphi' = -\lambda\,|\mathbf{a}| \cos(\pi - \varphi)$$
$$= -\lambda\,|\mathbf{a}|\,(-\cos\varphi) = \lambda\,|\mathbf{a}| \cos\varphi = \lambda\ \text{proj}_u\ \mathbf{a},$$

for $\lambda = 0$ equality (1) is obvious.

Thus, for any $\lambda$, $\text{proj}_u\ \lambda\mathbf{a} = \lambda\ \text{proj}_u\ \mathbf{a}$. ■

The theorems we have proved yield two important corollaries.

**Corollary 1.** *It follows from Theorem* 9.3 *that if* $\mathbf{a} = \{X_1, Y_1, Z_1\}$, $\mathbf{b} = \{X_2, Y_2, Z_2\}$, *then* $\mathbf{a} + \mathbf{b} = \{X_1 + X_2, Y_1 + Y_2, Z_1 + Z_2\}$.

**Corollary 2.** *It follows from Theorem* 9.4 *that if* $\mathbf{a} = \{X, Y, Z\}$, *then* $\lambda\mathbf{a} = \{\lambda X, \lambda Y, \lambda Z\}$ *for any scalar* $\lambda$.

From this corollary we can easily derive a condition for the collinearity of two vectors in coordinates. Indeed, the equality $\mathbf{b} = \lambda\mathbf{a}$

is equivalent to the equalities $X_2 = \lambda X_1$, $Y_2 = \lambda Y_1$, $Z_2 = \lambda Z_1$ or

$$\frac{X_2}{X_1} = \frac{Y_2}{Y_1} = \frac{Z_2}{Z_1}, \tag{2}$$

i.e. the vectors **a** and **b** are collinear if and only if their coordinates are proportional.

### 2.9.5. Representation of a Vector in Terms of Base Vectors

Let **i**, **j**, **k** be unit vectors of the coordinate axes, i.e. $|\mathbf{i}| = |\mathbf{j}| = |\mathbf{k}| = 1$, and each of them be of the same direction as the corresponding coordinate axis (Fig. 135). The triad of the vectors **i**, **j**, **k** is a *basis*.

Consider the following theorem.

**Theorem 9.5.** *Any vector* **a** *can be uniquely represented in terms of the base vectors* **i**, **j**, **k**, *i.e. in the form*

$$\mathbf{a} = \lambda\mathbf{i} + \mu\mathbf{j} + \nu\mathbf{k}, \tag{1}$$

*where* $\lambda$, $\mu$, $\nu$ *are some numbers.*

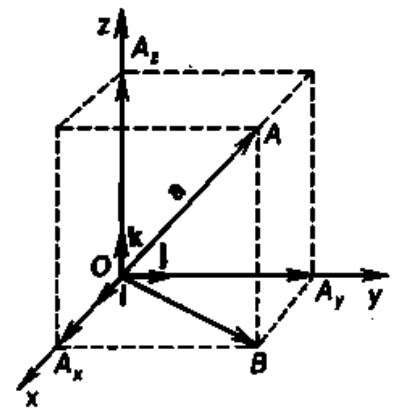


Fig. 135

*Proof.* Applying the vector **a** to the origin of coordinates, we designate its terminus as $A$. We draw planes perpendicular to the coordinate axes through the point $A$. Let $A_x$, $A_y$, $A_z$ be the points of intersection of the planes and the coordinate axes. By the definition of the addition of vectors, we have

$$\mathbf{a} = \overrightarrow{OB} + \overrightarrow{OA_z}, \quad \overrightarrow{OB} = \overrightarrow{OA_x} + \overrightarrow{OA_y}. \tag{2}$$

From relations (2) we obtain

$$\mathbf{a} = \overrightarrow{OA_x} + \overrightarrow{OA_y} + \overrightarrow{OA_z}. \tag{3}$$

Since the vectors $\overrightarrow{OA_x}$ and **i**, $\overrightarrow{OA_y}$ and **j**, $\overrightarrow{OA_z}$ and **k** are collinear, it follows that

$$\overrightarrow{OA_x} = \lambda\mathbf{i}, \quad \overrightarrow{OA_y} = \mu\mathbf{j}, \quad \overrightarrow{OA_z} = \nu\mathbf{k}, \tag{4}$$

where $\lambda$, $\mu$, $\nu$ are some numbers.

From relation (3) and relations (4) we find that

$$\mathbf{a} = \lambda\mathbf{i} + \mu\mathbf{j} + \nu\mathbf{k}.$$

To prove the uniqueness of representation (1), we establish that

$$\lambda = X, \quad \mu = Y, \quad \nu = Z,$$

where $X$, $Y$, $Z$ are the coordinates of the vector **a**.

We shall show, for example, that $\lambda = X$. Since $X = |\overrightarrow{OA_x}|$ if $\overrightarrow{OA_x}$ is of the same direction as the vector $\mathbf{i}$, and $X = -|\overrightarrow{OA_x}|$ if the vector $\overrightarrow{OA_x}$ is of the opposite direction to that of the vector $\mathbf{i}$, it follows that $\overrightarrow{OA_x} = X\mathbf{i}$. Comparing this relation with the relation $\overrightarrow{OA_x} = \lambda\mathbf{i}$, we get $\lambda = X$.

We can similarly prove that $\mu = Y$ and $\nu = Z$. ■

### 2.9.6. The Scalar Product of Vectors

**1. Definition of the scalar (dot) product.**

**Definition.** *The scalar product of two nonzero vectors* **a** *and* **b** *is a number (scalar) equal to the product of the lengths of those vectors by*

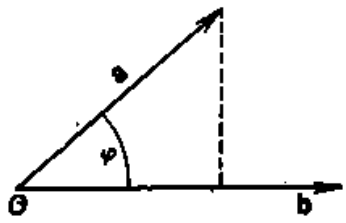



Fig. 136

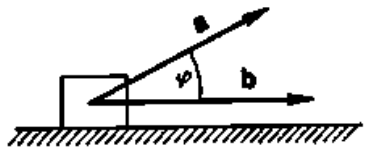



Fig. 137

*the cosine of the angle between them. If at least one of the vectors is zero, then the angle is not defined and, by definition, the scalar product is assumed to be zero.*

The scalar product of the vectors **a** and **b** is designated as $\mathbf{a}\cdot\mathbf{b}$. Thus

$$\mathbf{a}\cdot\mathbf{b} = |\mathbf{a}|\,|\mathbf{b}|\cos\varphi,$$

where $\varphi$ is the angle between the vectors **a** and **b** (Fig. 136).

Since $|\mathbf{a}|\cos\varphi = \text{proj}_{\mathbf{b}}\,\mathbf{a}$, $|\mathbf{b}|\cos\varphi = \text{proj}_{\mathbf{a}}\,\mathbf{b}$, we can write

$$\mathbf{a}\cdot\mathbf{b} = |\mathbf{b}|\,\text{proj}_{\mathbf{b}}\,\mathbf{a} = |\mathbf{a}|\,\text{proj}_{\mathbf{a}}\,\mathbf{b}. \qquad (1)$$

The formula for work

$$A = |\mathbf{a}|\,|\mathbf{b}|\cos\varphi$$

is a typical example of a scalar product. Here the vector **a** is the force whose point of application is displaced from the origin to the terminus of the vector **b** (Fig. 137).

**2. Principal properties of the scalar product.**

1°. $\mathbf{a}\cdot\mathbf{b} = \mathbf{b}\cdot\mathbf{a}$ (*commutativity of the factors*).

*Proof.* By the definition of a scalar product, $\mathbf{a}\cdot\mathbf{b} = |\mathbf{a}|\,|\mathbf{b}|\cos\varphi$ and $\mathbf{b}\cdot\mathbf{a} = |\mathbf{b}|\,|\mathbf{a}|\cos\varphi$, but $|\mathbf{a}|\,|\mathbf{b}| = |\mathbf{b}|\,|\mathbf{a}|$ since it is the product of numbers. Consequently, $\mathbf{a}\cdot\mathbf{b} = \mathbf{b}\cdot\mathbf{a}$. ■

2°. $(\lambda\mathbf{a})\cdot\mathbf{b} = \lambda(\mathbf{a}\cdot\mathbf{b})$ (*associativity relative to multiplication by a scalar*).

*Proof*. From formula (1) we find that

$$(\lambda\mathbf{a})\cdot\mathbf{b} = |\mathbf{b}|\,\mathrm{proj}_{\mathbf{b}}(\lambda\mathbf{a}),$$

but, according to Theorem 9.4,

$$\mathrm{proj}_{\mathbf{b}}(\lambda\mathbf{a}) = \lambda\,\mathrm{proj}_{\mathbf{b}}\,\mathbf{a}.$$

Thus

$$(\lambda\mathbf{a})\cdot\mathbf{b} = |\mathbf{b}|\,\mathrm{proj}_{\mathbf{b}}(\lambda\mathbf{a}) = |\mathbf{b}|\,\lambda\,\mathrm{proj}_{\mathbf{b}}\,\mathbf{a} = \lambda(|\mathbf{b}|\,\mathrm{proj}_{\mathbf{b}}\,\mathbf{a}).$$

On the other hand, from the same formula (1) we find that

$$|\mathbf{b}|\,\mathrm{proj}_{\mathbf{b}}\,\mathbf{a} = \mathbf{a}\cdot\mathbf{b}.$$

Consequently,

$$(\lambda\mathbf{a})\cdot\mathbf{b} = \lambda(|\mathbf{b}|\,\mathrm{proj}_{\mathbf{b}}\,\mathbf{a}) = \lambda(\mathbf{a}\cdot\mathbf{b}). \blacksquare$$

**Remark 1.** It follows from Properties 1° and 2° that $(\lambda\mathbf{a})\cdot(\mu\mathbf{b}) = (\lambda\cdot\mu)(\mathbf{a}\cdot\mathbf{b})$. Indeed, $(\lambda\mathbf{a})\cdot(\mu\mathbf{b}) = \lambda[\mathbf{a}\cdot(\mu\mathbf{b})] = \lambda[(\mu\mathbf{b})\cdot\mathbf{a}] = \lambda[\mu(\mathbf{b}\cdot\mathbf{a})] = (\lambda\cdot\mu)(\mathbf{a}\cdot\mathbf{b})$.

3°. $\mathbf{a}\cdot(\mathbf{b}+\mathbf{c}) = \mathbf{a}\cdot\mathbf{b} + \mathbf{a}\cdot\mathbf{c}$ (*distributivity relative to the sum of vectors*).

*Proof*. From formula (1) we find that

$$\mathbf{a}\cdot(\mathbf{b}+\mathbf{c}) = |\mathbf{a}|\,\mathrm{proj}_{\mathbf{a}}(\mathbf{b}+\mathbf{c}),$$

but, according to Theorem 9.3,

$$\mathrm{proj}_{\mathbf{a}}(\mathbf{b}+\mathbf{c}) = \mathrm{proj}_{\mathbf{a}}\,\mathbf{b} + \mathrm{proj}_{\mathbf{a}}\,\mathbf{c}.$$

Thus

$$\begin{aligned}\mathbf{a}\cdot(\mathbf{b}+\mathbf{c}) &= |\mathbf{a}|\,\mathrm{proj}_{\mathbf{a}}(\mathbf{b}+\mathbf{c}) = |\mathbf{a}|(\mathrm{proj}_{\mathbf{a}}\,\mathbf{b} + \mathrm{proj}_{\mathbf{a}}\,\mathbf{c})\\ &= |\mathbf{a}|\,\mathrm{proj}_{\mathbf{a}}\,\mathbf{b} + |\mathbf{a}|\,\mathrm{proj}_{\mathbf{a}}\,\mathbf{c}.\end{aligned}$$

On the other hand, from the same formula (1) we get

$$|\mathbf{a}|\,\mathrm{proj}_{\mathbf{a}}\,\mathbf{b} = \mathbf{a}\cdot\mathbf{b}, \quad |\mathbf{a}|\,\mathrm{proj}_{\mathbf{a}}\,\mathbf{c} = \mathbf{a}\cdot\mathbf{c}.$$

Consequently,

$$\mathbf{a}\cdot(\mathbf{b}+\mathbf{c}) = |\mathbf{a}|\,\mathrm{proj}_{\mathbf{a}}\,\mathbf{b} + |\mathbf{a}|\,\mathrm{proj}_{\mathbf{a}}\,\mathbf{c} = \mathbf{a}\cdot\mathbf{b} + \mathbf{a}\cdot\mathbf{c}. \blacksquare$$

**Remark 2.** The property we have proved makes it legitimate to perform the actions term by term when carrying out a scalar multiplication of vector polynomials. In that case, by virtue of Property 1°, we can ignore the sequence of factors, and Property 2° (see Remark 1) allows us to combine the numerical coefficients of the vector factors. For instance,

$$\begin{aligned}(2\mathbf{a}+5\mathbf{b})\cdot(3\mathbf{c}+4\mathbf{d}) &= (2\mathbf{a}+5\mathbf{b})\cdot(3\mathbf{c}) + (2\mathbf{a}+5\mathbf{b})\cdot(4\mathbf{d})\\ &= (2\mathbf{a})\cdot(3\mathbf{c}) + (5\mathbf{b})\cdot(3\mathbf{c}) + (2\mathbf{a})\cdot(4\mathbf{d}) + (5\mathbf{b})\cdot(4\mathbf{d})\\ &= 6\mathbf{a}\cdot\mathbf{c} + 15\mathbf{b}\cdot\mathbf{c} + 8\mathbf{a}\cdot\mathbf{d} + 20\mathbf{b}\cdot\mathbf{d}.\end{aligned}$$

4°. $\mathbf{a}\cdot\mathbf{a} = |\mathbf{a}|^2$.

*Proof.* By the definition of a scalar product, $\mathbf{a}\cdot\mathbf{a} = |\mathbf{a}|\,|\mathbf{a}|\cos 0 = |\mathbf{a}|^2$ if $|\mathbf{a}| \neq 0$, i.e. if $\mathbf{a} \neq \mathbf{0}$. Now if $\mathbf{a} = \mathbf{0}$, then, again by definition, $\mathbf{a}\cdot\mathbf{a} = 0$. But in this case $|\mathbf{a}| = 0$, and this means that the equality $\mathbf{a}\cdot\mathbf{a} = |\mathbf{a}|^2$ also holds true. ■

The scalar product $\mathbf{a}\cdot\mathbf{a}$ is known as the *dot product* of the vector $\mathbf{a}$ with itself and is designated as $\mathbf{a}^2$. By virtue of what we have just proved, we have $\mathbf{a}^2 = |\mathbf{a}|^2$. From this it follows, in particular, that $\sqrt{\mathbf{a}^2} = |\mathbf{a}|$.

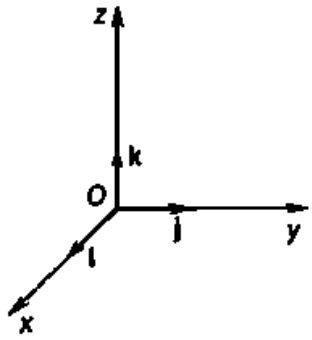


Fig. 138

5°. $\mathbf{a}\cdot\mathbf{b} = 0$ if $\mathbf{a} \perp \mathbf{b}$ and, conversely, $\mathbf{a} \perp \mathbf{b}$ if $\mathbf{a}\cdot\mathbf{b} = 0$ and $\mathbf{a} \neq \mathbf{0}$, $\mathbf{b} \neq \mathbf{0}$.

*Proof.* By the definition of the scalar product, $\mathbf{a}\cdot\mathbf{b} = |\mathbf{a}|\,|\mathbf{b}|\cos\varphi$. If $\varphi = \pi/2$, i.e. if the vectors $\mathbf{a}$ and $\mathbf{b}$ are perpendicular to each other, then $\cos\varphi = 0$, and, hence, $\mathbf{a}\cdot\mathbf{b} = 0$.

Conversely, if $\mathbf{a}\cdot\mathbf{b} = 0$ and $|\mathbf{a}|\,|\mathbf{b}| \neq 0$, then $\cos\varphi = 0$ and $\varphi = \pi/2$, i.e. the vectors $\mathbf{a}$ and $\mathbf{b}$ are perpendicular to each other. ■

**Remark 3.** From Properties 4° and 5° of base vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$ (Fig. 138) we immediately obtain the following equalities:

$$\begin{gathered} \mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = 1, \\ \mathbf{i}\cdot\mathbf{j} = \mathbf{i}\cdot\mathbf{k} = \mathbf{j}\cdot\mathbf{i} = \mathbf{j}\cdot\mathbf{k} = \mathbf{k}\cdot\mathbf{i} = \mathbf{k}\cdot\mathbf{j} = 0. \end{gathered} \tag{2}$$

**3. Expressing a scalar product in terms of the coordinates of the vectors.**

**Theorem 9.6.** *If the vectors* $\mathbf{a}$ *and* $\mathbf{b}$ *are defined by their coordinates, i.e.* $\mathbf{a} = \{X_1, Y_1, Z_1\}$, $\mathbf{b} = \{X_2, Y_2, Z_2\}$, *then their scalar product is defined by the formula*

$$\mathbf{a}\cdot\mathbf{b} = X_1X_2 + Y_1Y_2 + Z_1Z_2.$$

*Proof.* We resolve the vectors $\mathbf{a}$ and $\mathbf{b}$ into components with respect to the base vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$: $\mathbf{a} = X_1\mathbf{i} + Y_1\mathbf{j} + Z_1\mathbf{k}$, $\mathbf{b} = X_2\mathbf{i} + Y_2\mathbf{j} + Z_2\mathbf{k}$. Using Remark 2, we obtain

$$\begin{aligned} \mathbf{a}\cdot\mathbf{b} = {} & X_1X_2\mathbf{i}^2 + X_1Y_2\mathbf{i}\cdot\mathbf{j} + X_1Z_2\mathbf{i}\cdot\mathbf{k} \\ & + Y_1X_2\mathbf{j}\cdot\mathbf{i} + Y_1Y_2\mathbf{j}^2 + Y_1Z_2\mathbf{j}\cdot\mathbf{k} + Z_1X_2\mathbf{k}\cdot\mathbf{i} \\ & + Z_1Y_2\mathbf{k}\cdot\mathbf{j} + Z_1Z_2\mathbf{k}^2, \end{aligned}$$

whence, using relations (2), we find that $\mathbf{a}\cdot\mathbf{b} = X_1X_2 + Y_1Y_2 + Z_1Z_2$. ■

Theorem 9.6 yields two important corollaries.

**Corollary 1.** *The necessary and sufficient condition for the perpendicularity of the vectors* $\mathbf{a} = \{X_1, Y_1, Z_1\}$ *and* $\mathbf{b} = \{X_2, Y_2, Z_2\}$ *is the relation*

$$X_1X_2 + Y_1Y_2 + Z_1Z_2 = 0. \tag{3}$$

This statement follows directly from Property 5° and Theorem 9.6.

**Corollary 2.** *The angle between the vectors* $\mathbf{a} = \{X_1, Y_1, Z_1\}$ *and* $\mathbf{b} = \{X_2, Y_2, Z_2\}$ *is defined by the relation*

$$\cos\varphi = \frac{X_1X_2+Y_1Y_2+Z_1Z_2}{\sqrt{X_1^2+Y_1^2+Z_1^2}\sqrt{X_2^2+Y_2^2+Z_2^2}}. \tag{4}$$

Indeed, by the definition of the scalar product, $\mathbf{a}\cdot\mathbf{b} = |\mathbf{a}|\,|\mathbf{b}|\cos\varphi$, whence it follows that

$$\cos\varphi = \frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}|\,|\mathbf{b}|}. \tag{5}$$

By virtue of Theorem 9.6 and formula (4) of 2.9.2 relation (5) yields relation (4).

**Example.** Given three points $A$ (1, 1, 1), $B$ (2, 2, 1), $C$ (2, 1, 2), find the angle $\varphi = \angle BAC$.

*Solution.* Using Theorem 9.2, we find that $\overrightarrow{AB} = \{1, 1, 0\}$, $\overrightarrow{AC} = \{1, 0, 1\}$. From this, on the basis of relation (4), we get

$$\cos\varphi = \frac{1\cdot1+1\cdot0+0\cdot1}{\sqrt{1^2+1^2+0^2}\sqrt{1^2+0^2+1^2}} = \frac{1}{\sqrt{2}\sqrt{2}} = \frac{1}{2}.$$

Consequently, $\varphi = 60°$.

### 2.9.7. The Vector Product of Vectors

**1. Definition of the vector (cross) product.** The vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ are *coplanar* if they lie in the same plane or in parallel planes.

The triad of vectors is said to be *ordered* (a trihedral of vectors) if it is indicated which of them is the first, which is the second and which is the third. For instance, in the notation ($\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$) the vector $\mathbf{a}$ is considered to be the first, $\mathbf{b}$ the second and $\mathbf{c}$ the third; in the notation ($\mathbf{b}$, $\mathbf{c}$, $\mathbf{a}$) the vector $\mathbf{b}$ is the first, $\mathbf{c}$ the second and $\mathbf{a}$ the third.

A trihedral of noncoplanar vectors is *right-handed* if, when they are drawn from the common point, the shortest rotation from the first vector to the second is seen to be in the counterclockwise direction when we contemplate the rotation from the terminus of the third vector. Otherwise the trihedral of vectors is *left-handed.*

**Definition.** *The vector (cross) product of the vector* $\mathbf{a}$ *by the vector* $\mathbf{b}$ *is a vector* $\mathbf{a}\times\mathbf{b}$ *which is defined by three conditions:*

(1) *the length of the vector* $\mathbf{a}\times\mathbf{b}$ *is equal to* $|\mathbf{a}|\,|\mathbf{b}|\sin\varphi$, *where* $\varphi$ *is the angle between the vectors* $\mathbf{a}$ *and* $\mathbf{b}$;

(2) *the vector* $\mathbf{a}\times\mathbf{b}$ *is perpendicular to each of the vectors* $\mathbf{a}$ *and* $\mathbf{b}$;

(3) *the vectors* $\mathbf{a}$, $\mathbf{b}$, $\mathbf{a}\times\mathbf{b}$ *form a right-handed trihedral* (Fig. 139).

Note that conditions (2) and (3) refer to the case when $|\mathbf{a}|\,|\mathbf{b}|\times\sin\varphi \neq 0$, i.e. the vector $\mathbf{a}\times\mathbf{b} \neq 0$. Now if $|\mathbf{a}|\,|\mathbf{b}|\sin\varphi = 0$

(i.e. either at least one of the vectors **a** and **b** is zero or $\sin\varphi = 0$), the vector product $\mathbf{a} \times \mathbf{b}$ is defined only by condition (1). Here $\mathbf{a} \times \mathbf{b} = 0$. The concept of a vector product originated in mechanics.

Assume that a force $\mathbf{F} = \overrightarrow{MK}$ is applied at a point $M$ of a solid body and $O$ is a point of space. It is known from mechanics that the moment of force **F** relative to the point $O$ (the point of application of the moment) is the vector **L** which (1) has a length equal to $|\overrightarrow{OM}|\,|\overrightarrow{MK}|\sin\varphi$,

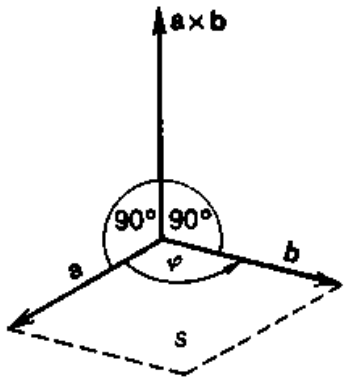



Fig. 139

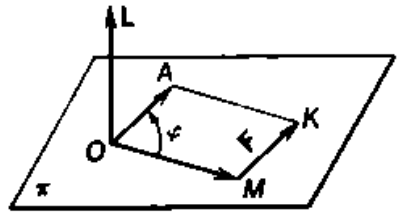



Fig. 140

where $\varphi$ is the angle between the vectors $\overrightarrow{OM}$ and $\overrightarrow{MK}$; (2) is perpendicular to the plane $\pi$ which passes through the points $O$, $M$, $K$; (3) is directed so that when contemplated from its terminus, the force **F** seems to rotate the plane $\pi$ about the point $O$ counterclockwise (Fig. 140). From the figure in which $\overrightarrow{OA} = \overrightarrow{MK}$ we can see that **L** is the vector product $\overrightarrow{OM} \times \overrightarrow{MK}$.

**2. Principal properties of the vector product.**

1°. $\mathbf{a} \times \mathbf{b} = 0$ *if* **a** *and* **b** *are collinear vectors.*

*Proof.* If the vectors **a** and **b** are collinear, then $\sin\varphi = 0$. Consequently, $|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}|\,|\mathbf{b}|\sin\varphi = 0$, i.e. the length of the vector $\mathbf{a} \times \mathbf{b}$ is zero, and, hence, the vector $\mathbf{a} \times \mathbf{b}$ itself is zero. ■

2°. *The length of the vector product of the noncollinear vectors* **a** *and* **b** *is equal to the area s of the parallelogram constructed on these vectors* (see Fig. 139).

*Proof.* It is known from elementary geometry that the area of a parallelogram is equal to the product of its adjacent sides by the sine of the angle between them. Hence $|\mathbf{a}|\,|\mathbf{b}|\sin\varphi = s$, i.e. $|\mathbf{a} \times \mathbf{b}| = s$. ■

3°. $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$ (*anticommutativity of the factors*).

*Proof.* If the vectors **a** and **b** are collinear, then the property is obvious. Let **a** and **b** be noncollinear. From the definition of the vector product it follows that the vectors $\mathbf{a} \times \mathbf{b}$ and $\mathbf{b} \times \mathbf{a}$ are of the same length (the length of a vector product does not depend on the order of the factors), are collinear (they are perpendicular to the same plane in which the vectors **a** and **b** lie) but are opposite in

direction (Fig. 141) since the vectors **a**, **b**, **a** × **b** and **b**, **a**, **b** × **a** form right-handed trihedrals. Consequently, **a** × **b** = −**b** × **a**. ■

4°. (λ**a**) × **b** = λ (**a** × **b**) (*associativity relative to the scalar factor*).

*Proof.* If **a** and **b** are collinear or λ = 0, then the property is obvious. Let **a** and **b** be noncollinear and λ ≠ 0. From the definition of the vector product we have | λ (**a** × **b**) | = | λ | | **a** | | **b** | sin φ and | (λ**a**) × **b** | = | λ | | **a** | | **b** | sin φ, and therefore the vec-

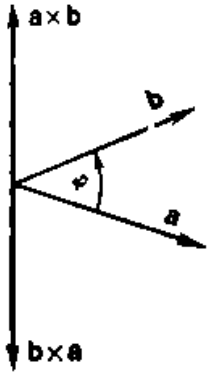



Fig. 141

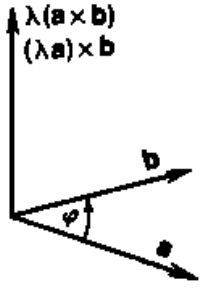



Fig. 142

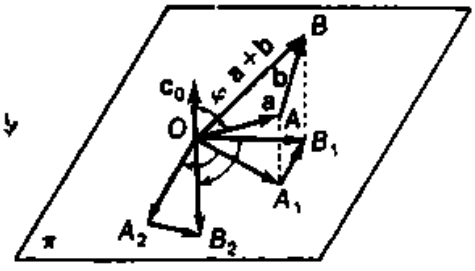



Fig. 143

tors (λ**a**) × **b** and λ (**a** × **b**) are of the same length. In addition, they are perpendicular to each of the vectors **a** and **b** and, hence, are collinear. Finally, they are of the same direction (Fig. 142) (for λ > 0 this is obvious since the vectors λ**a** and **a** are of the same direction; for λ < 0 the vectors λ**a** and **a** are of opposite directions and therefore the vector (λ**a**) × **b** is of the direction opposite to that of the vector **a** × **b**, but in this case the direction of the vector λ (**a** × **b**) is also opposite to that of the vector **a** × **b** and, hence, the vectors (λ**a**) × **b** and λ (**a** × **b**) have the same direction for λ < 0 as well). Consequently, the vectors (λ**a**) × **b** and λ (**a** × **b**) are equal. ■

Using Properties 3° and 4°, prove that **a** × (λ**b**) = λ (**a** × **b**).

5°. (**a** + **b**) × **c** = **a** × **c** + **b** × **c** (*distributivity relative to the sum of vectors*).

*Proof.* If the vectors **a** and **b** are collinear with the vector **c** or at least one of the vectors **a**, **b**, **c** is zero, then the property is obvious. In other cases we introduce a unit vector $\mathbf{c}_0$ which has the same direction as the vector **c** and lies on the same straight line. Through its origin $O$ we draw a plane π at right angles to $\mathbf{c}_0$ and consider a triangle $OAB$ such that $\overrightarrow{OA} = \mathbf{a}$, $\overrightarrow{AB} = \mathbf{b}$, $\overrightarrow{OB} = \mathbf{a} + \mathbf{b}$ (Fig. 143). We project the triangle $OAB$ onto the plane π and obtain a triangle $OA_1B_1$ (if the point $A_1$ lies on the straight line $OB_1$, then the triangle $OA_1B_1$ degenerates into a line segment). We rotate the triangle $OA_1B_1$ about $\mathbf{c}_0$ through 90° clockwise if we contemplate it from the terminus of $\mathbf{c}_0$, and obtain a triangle $OA_2B_2$. We designate the angle between the vectors $\mathbf{c}_0$ and **a** as φ. Let, for definiteness, $0 < \varphi < \pi/2$.

The other cases of the angle $\varphi$ can be treated by analogy.

Let us consider the vector $\overrightarrow{OA_2}$. Its length $|\overrightarrow{OA_2}| = |\overrightarrow{OA_1}| = |\mathbf{a}| \cos(\pi/2 - \varphi) = |\mathbf{a}|\,|\mathbf{c}_0| \sin\varphi$ since $|\mathbf{c}_0| = 1$. In addition, $\overrightarrow{OA_2} \perp \mathbf{c}_0$, $\overrightarrow{OA_2} \perp \mathbf{a}$ and the vectors $\mathbf{a}$, $\mathbf{c}_0$, $\overrightarrow{OA_2}$ form a right-handed trihedral. Consequently, by the definition of the vector product,

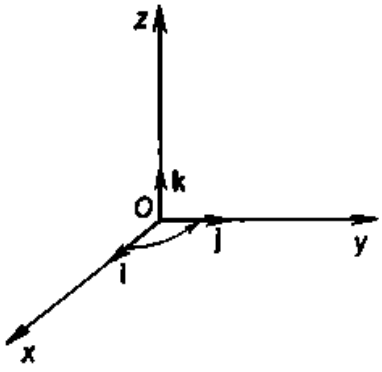


Fig. 1

$$\overrightarrow{OA_2} = \mathbf{a} \times \mathbf{c}_0.$$

Reasoning by analogy for each of the vectors $\overrightarrow{OB_2}$ and $\overrightarrow{A_2B_2}$, we obtain

$$\overrightarrow{OB_2} = (\mathbf{a} + \mathbf{b}) \times \mathbf{c}_0, \quad \overrightarrow{A_2B_2} = \mathbf{b} \times \mathbf{c}_0.$$

But since $\overrightarrow{OB_2} = \overrightarrow{OA_2} + \overrightarrow{A_2B_2}$, it follows that

$$(\mathbf{a} + \mathbf{b}) \times \mathbf{c}_0 = \mathbf{a} \times \mathbf{c}_0 + \mathbf{b} \times \mathbf{c}_0. \quad (1)$$

The direction of the vector $\mathbf{c}$ is the same as of $\mathbf{c}_0$ and therefore $\mathbf{c} = |\mathbf{c}|\,\mathbf{c}_0$. Multiplying both sides of relation (1) by the number $|\mathbf{c}|$, we get $|\mathbf{c}|\,[(\mathbf{a} + \mathbf{b}) \times \mathbf{c}_0] = |\mathbf{c}|\,(\mathbf{a} \times \mathbf{c}_0 + \mathbf{b} \times \mathbf{c}_0)$. From this, according to Property 4°, it follows that $(\mathbf{a} + \mathbf{b}) \times |\mathbf{c}|\,\mathbf{c}_0 = \mathbf{a} \times |\mathbf{c}|\,\mathbf{c}_0 + \mathbf{b} \times |\mathbf{c}|\,\mathbf{c}_0$. Replacing $|\mathbf{c}|\,\mathbf{c}_0$ by $\mathbf{c}$, we finally obtain $(\mathbf{a} + \mathbf{b}) \times \mathbf{c} = \mathbf{a} \times \mathbf{c} + \mathbf{b} \times \mathbf{c}$. ■

**Remark 1.** The property we have proved makes it legitimate to carry out the operations term by term when performing a vector multiplication of vector polynomials, and Property 4°, to combine the numerical coefficients of the vector factors. For example,

$$\begin{aligned}(2\mathbf{a} + 3\mathbf{b}) \times (4\mathbf{c} + 5\mathbf{d}) &= (2\mathbf{a} + 3\mathbf{b}) \times 4\mathbf{c} + (2\mathbf{a} + 3\mathbf{b}) \times 5\mathbf{d} \\ &= 2\mathbf{a} \times 4\mathbf{c} + 3\mathbf{b} \times 4\mathbf{c} + 2\mathbf{a} \times 5\mathbf{d} + 3\mathbf{b} \times 5\mathbf{d} \\ &= 8\,(\mathbf{a} \times \mathbf{c}) + 12\,(\mathbf{b} \times \mathbf{c}) + 10\,(\mathbf{a} \times \mathbf{d}) + 15\,(\mathbf{b} \times \mathbf{d}).\end{aligned}$$

It should be remembered, however, that the sequence of the factors in a vector product is essential and so the sign of the vector product must be changed when the factors are interchanged. For example, $(\mathbf{a} + \mathbf{b} + \mathbf{c}) \times (2\mathbf{b} + 3\mathbf{c}) = 2\,(\mathbf{a} \times \mathbf{b}) + 2\,(\mathbf{b} \times \mathbf{b}) + 2\,(\mathbf{c} \times \mathbf{b}) + 3\,(\mathbf{a} \times \mathbf{c}) + 3\,(\mathbf{b} \times \mathbf{c}) + 3\,(\mathbf{c} \times \mathbf{c}) = 2\,(\mathbf{a} \times \mathbf{b}) - 2\,(\mathbf{b} \times \mathbf{c}) + 3\,(\mathbf{a} \times \mathbf{c}) + 3\,(\mathbf{b} \times \mathbf{c}) = 2\,(\mathbf{a} \times \mathbf{b}) + (\mathbf{b} \times \mathbf{c}) + 3\,(\mathbf{a} \times \mathbf{c})$.

**Remark 2.** According to the definition and to Properties 1° and 3° of the vector product, we get the following relations for the base vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ (Fig. 144):

$$\begin{aligned}&\mathbf{i} \times \mathbf{i} = 0, \quad \mathbf{i} \times \mathbf{j} = \mathbf{k}, \quad \mathbf{i} \times \mathbf{k} = -\mathbf{j}, \\ &\mathbf{j} \times \mathbf{i} = -\mathbf{k}, \quad \mathbf{j} \times \mathbf{j} = 0, \quad \mathbf{j} \times \mathbf{k} = \mathbf{i}, \\ &\mathbf{k} \times \mathbf{i} = \mathbf{j}, \quad \mathbf{k} \times \mathbf{j} = -\mathbf{i}, \quad \mathbf{k} \times \mathbf{k} = 0.\end{aligned} \quad (2)$$

**3. Expressing a vector product in terms of the coordinates of the vectors.**

**Theorem 9.7.** *If the vectors* **a** *and* **b** *are defined by their coordinates, i.e.* $\mathbf{a} = \{X_1, Y_1, Z_1\}$ *and* $\mathbf{b} = \{X_2, Y_2, Z_2\}$, *then the vector product of the vectors* **a** *and* **b** *is defined by the formula*

$$\mathbf{a} \times \mathbf{b} = \{(Y_1Z_2 - Y_2Z_1),\ (Z_1X_2 - Z_2X_1),\ (X_1Y_2 - X_2Y_1)\}.$$

Using second-order determinants*, we can write this formula as

$$\mathbf{a} \times \mathbf{b} = \left\{ \begin{vmatrix} Y_1 & Z_1 \\ Y_2 & Z_2 \end{vmatrix}, \begin{vmatrix} Z_1 & X_1 \\ Z_2 & X_2 \end{vmatrix}, \begin{vmatrix} X_1 & Y_1 \\ X_2 & Y_2 \end{vmatrix} \right\}.$$

*Proof.* We resolve the vectors **a** and **b** into components with respect to the base vectors **i**, **j**, **k**:

$$\mathbf{a} = X_1\mathbf{i} + Y_1\mathbf{j} + Z_1\mathbf{k}, \quad \mathbf{b} = X_2\mathbf{i} + Y_2\mathbf{j} + Z_2\mathbf{k}.$$

Using Remark 1, we obtain

$$\begin{aligned} \mathbf{a} \times \mathbf{b} = {} & X_1X_2(\mathbf{i} \times \mathbf{i}) + X_1Y_2(\mathbf{i} \times \mathbf{j}) + X_1Z_2(\mathbf{i} \times \mathbf{k}) \\ & + Y_1X_2(\mathbf{j} \times \mathbf{i}) + Y_1Y_2(\mathbf{j} \times \mathbf{j}) + Y_1Z_2(\mathbf{j} \times \mathbf{k}) \\ & + Z_1X_2(\mathbf{k} \times \mathbf{i}) + Z_1Y_2(\mathbf{k} \times \mathbf{j}) + Z_1Z_2(\mathbf{k} \times \mathbf{k}). \end{aligned}$$

From this, on the basis of relations (2), we find that

$$\mathbf{a} \times \mathbf{b} = (Y_1Z_2 - Y_2Z_1)\,\mathbf{i} + (Z_1X_2 - Z_2X_1)\,\mathbf{j} + (X_1Y_2 - X_2Y_1)\,\mathbf{k},$$

or

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} Y_1 & Z_1 \\ Y_2 & Y_2 \end{vmatrix}\mathbf{i} + \begin{vmatrix} Z_1 & X_1 \\ Z_2 & X_2 \end{vmatrix}\mathbf{j} + \begin{vmatrix} X_1 & Y_1 \\ X_2 & Y_2 \end{vmatrix}\mathbf{k}.$$

We have obtained the representation of the vector $\mathbf{a} \times \mathbf{b}$ in terms of the base vectors **i**, **j**, **k**. The coefficients in this representation are the coordinates of the vector $\mathbf{a} \times \mathbf{b}$. Thus we have

$$\mathbf{a} \times \mathbf{b} = \{X, Y, Z\}, \tag{3}$$

where

$$X = \begin{vmatrix} Y_1 & Z_1 \\ Y_2 & Z_2 \end{vmatrix}, \quad Y = \begin{vmatrix} Z_1 & X_1 \\ Z_2 & X_2 \end{vmatrix}, \quad Z = \begin{vmatrix} X_1 & Y_1 \\ X_2 & Y_2 \end{vmatrix}. \ \blacksquare$$

**Example.** Given vectors $\mathbf{a} = \{2, 5, 7\}$ and $\mathbf{b} = \{1, 2, 4\}$, find the coordinates of the vector product $\mathbf{a} \times \mathbf{b}$.

*Solution.* From formula (3) we find that

$$X = \begin{vmatrix} 5 & 7 \\ 2 & 4 \end{vmatrix} = 6, \quad Y = \begin{vmatrix} 7 & 2 \\ 4 & 1 \end{vmatrix} = -1, \quad Z = \begin{vmatrix} 2 & 5 \\ 1 & 2 \end{vmatrix} = -1.$$

Thus $\mathbf{a} \times \mathbf{b} = \{6, -1, -1\}$.

* A second-order determinant $\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$ is a number equal to $a_1b_2 - a_2b_1$.

## 2.9.8. The Triple Product of Three Vectors

**1. Definition and geometric meaning of a triple product.**

**Definition.** *The triple, or parallelepipedal, product of three vectors* **a**, **b**, **c** *is a number equal to the scalar product of the vector* **a** *by the vector product of the vectors* **b** *and* **c**, *i.e.*

$$\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}).$$

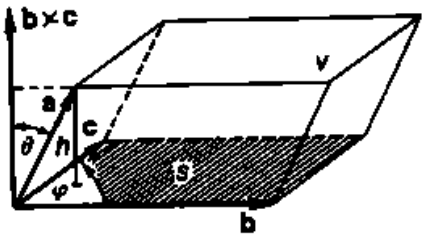


Fig. 145

The following theorem expresses the geometric meaning of a triple product.

**Theorem 9.8.** *The triple product* $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})$ *is equal to the volume* $v$ *of the parallelepiped constructed on the vectors* **a**, **b**, **c** *and taken with the plus sign if the triad* **a**, **b**, **c** *is right-handed and with the minus sign if the triad* **a**, **b**, **c** *is left-handed. Now if the vectors* **a**, **b**, **c** *are coplanar, then* $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = 0$. *In other words,*

$$\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = \begin{cases} v \text{ if } \mathbf{a}, \mathbf{b}, \mathbf{c} \text{ is a right-handed triad,} \\ -v \text{ if } \mathbf{a}, \mathbf{b}, \mathbf{c} \text{ is a left-handed triad,} \\ 0 \text{ if } \mathbf{a}, \mathbf{b}, \mathbf{c} \text{ are coplanar.} \end{cases}$$

*Proof.* Let **a**, **b**, **c** be noncoplanar vectors which form a right-handed triad. We designate the volume of the parallelepiped constructed on these vectors as $v$, the area of the parallelogram constructed on the vectors **b** and **c** as $s$, and the altitude of the parallelepiped as $h$ (Fig. 145). Then, by the definitions of the scalar and the vector product, we have

$$\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = |\mathbf{a}|\,|\mathbf{b}\times\mathbf{c}|\cos\theta = |\mathbf{a}|\,|\mathbf{b}|\,|\mathbf{c}|\sin\varphi\cos\theta$$
$$= |\mathbf{b}|\,|\mathbf{c}|\sin\varphi\,|\mathbf{a}|\cos\theta,$$

where $\varphi$ is the angle between the vectors **b** and **c**, and $\theta$ is the angle between the vectors **a** and $\mathbf{b}\times\mathbf{c}$. Since $|\mathbf{b}|\,|\mathbf{c}|\sin\varphi = s$, $|\mathbf{a}|\cos\theta = h$, it follows that $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = sh = v$. If the triad **a**, **b**, **c** is left-handed, then $h = |\mathbf{a}|\cos(\pi-\theta) = -|\mathbf{a}|\cos\theta$. Therefore, $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = -sh = -v$. We have proved the first statement of the theorem.

Let us prove the second statement. Let the vectors **a**, **b**, **c** be coplanar. If $\mathbf{a} = 0$, then, evidently, $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = 0$. Let $\mathbf{a} \neq 0$. Then either $\mathbf{b}\times\mathbf{c} = 0$ (if the vectors **b** and **c** are collinear) or $(\mathbf{b}\times\mathbf{c}) \perp \mathbf{a}$ (if **b** and **c** are noncollinear). In either case $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = 0$. ■

We have thus proved that if the vectors **a**, **b**, **c** are coplanar, then $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = 0$. The converse is also true: if $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = 0$, then the vectors **a**, **b**, **c** are coplanar. Indeed, were the vectors **a**, **b**, **c**

noncoplanar, then, by Theorem 9.8, the triple product $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = \pm v \neq 0$, and this contradicts the hypothesis.

**Corollary.** *The theorem easily yields the following identity:*

$$\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}, \tag{1}$$

*i.e. in a triple product the signs · and × can be interchanged.*

Indeed, according to Property 1° of a scalar product,

$$(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c} = \mathbf{c}\cdot(\mathbf{a}\times\mathbf{b}). \tag{2}$$

Furthermore, by Theorem 9.8 we have

$$\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = \pm v, \quad \mathbf{c}\cdot(\mathbf{a}\times\mathbf{b}) = \pm v. \tag{3}$$

Since the triads **a**, **b**, **c** and **c**, **a**, **b** have the same orientation, i.e. are either both right-handed or both left-handed, we must take the same sign on the right-hand sides of relations (3) on the basis of Theorem 9.8. Thus we have

$$\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = \mathbf{c}\cdot(\mathbf{a}\times\mathbf{b})$$

and, on the basis of relations, (2) we have

$$\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\times\mathbf{b})\cdot\mathbf{c},$$

i.e. we get identity (1).

By virtue of identity (1), the triple products $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})$ and $(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}$ can be denoted by a simple symbol **abc**.

**2. Representation of a triple product in terms of the coordinates of the vectors.**

**Theorem 9.9.** *If the vectors* **a**, **b**, **c** *are defined by their coordinates*

$$\mathbf{a} = \{X_1, Y_1, Z_1\}, \quad \mathbf{b} = \{X_2, Y_2, Z_2\}, \quad \mathbf{c} = \{X_3, Y_3, Z_3\},$$

*then the triple product* **abc** *is defined by the formula*

$$\mathbf{abc} = X_1\begin{vmatrix} Y_2 & Z_2 \\ Y_3 & Z_3 \end{vmatrix} + Y_1\begin{vmatrix} Z_2 & X_2 \\ Z_3 & X_3 \end{vmatrix} + Z_1\begin{vmatrix} X_2 & Y_2 \\ X_3 & Y_3 \end{vmatrix}.$$

*Proof.* We have $\mathbf{abc} = \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})$. According to Theorem 9.7,

$$\mathbf{b}\times\mathbf{c} = \left\{\begin{vmatrix} Y_2 & Z_2 \\ Y_3 & Z_3 \end{vmatrix}, \begin{vmatrix} Z_2 & X_2 \\ Z_3 & X_3 \end{vmatrix}, \begin{vmatrix} X_2 & Y_2 \\ X_3 & Y_3 \end{vmatrix}\right\}.$$

Performing a scalar multiplication of the vector $\mathbf{a} = \{X_1, Y_1, Z_1\}$ by the vector $\mathbf{b}\times\mathbf{c}$ and using Theorem 9.6, we obtain

$$\mathbf{abc} = X_1\begin{vmatrix} Y_2 & Z_2 \\ Y_3 & Z_3 \end{vmatrix} + Y_1\begin{vmatrix} Z_2 & X_2 \\ Z_3 & X_3 \end{vmatrix} + Z_1\begin{vmatrix} X_2 & Y_2 \\ X_3 & Y_3 \end{vmatrix}. \quad \blacksquare$$

**Example.** Given four points $A$ (1, 1, 1), $B$ (4, 4, 4), $C$ (3, 5, 5), $D$ (2, 4, 7) in space, find the volume of the tetrahedron $ABCD$.

*Solution.* As is known from elementary geometry, the volume $v_t$ of the tetrahedron $ABCD$ is equal to one-sixth of the volume of the parallelepiped constructed on the vectors $\overrightarrow{AB}$, $\overrightarrow{AC}$, $\overrightarrow{AD}$. From this and from Theorem 9.8 we infer that the volume $v_t$ is equal to 1/6 of the absolute value of the triple product $\overrightarrow{AB}\cdot\overrightarrow{AC}\cdot\overrightarrow{AD}$. We seek this triple product. We first find the coordinates of the vectors $\overrightarrow{AB}$, $\overrightarrow{AC}$, $\overrightarrow{AD}$. From Theorem 9.2 we have: $\overrightarrow{AB} = \{3, 3, 3\}$, $\overrightarrow{AC} = \{2, 4, 4\}$, $\overrightarrow{AD} = \{1, 3, 6\}$. Using Theorem 9.8, we get

$$\overrightarrow{AB}\cdot\overrightarrow{AC}\cdot\overrightarrow{AD}=3\begin{vmatrix}4 & 4\\ 3 & 6\end{vmatrix}+3\begin{vmatrix}4 & 2\\ 6 & 1\end{vmatrix}+3\begin{vmatrix}2 & 4\\ 1 & 3\end{vmatrix}=3\cdot 12-3\cdot 8+3\cdot 2=18.$$

Hence

$$v_t=\frac{1}{6}\cdot 18=3.$$

### 2.9.9. Equations of a Surface and of a Curve

Assume that we are given a rectangular system of coordinates $Oxyz$, an arbitrary surface $S$ (Fig. 146) and an equation

$$F(x, y, z) = 0. \tag{1}$$

*We say that equation* (1) *is an equation of the surface $S$ in the given system of coordinates if it is satisfied by the coordinates of any point*

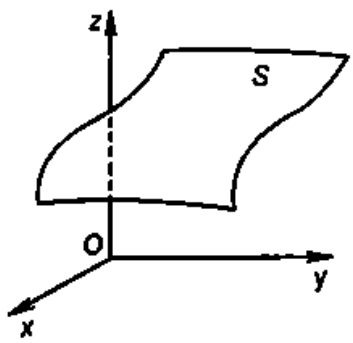



Fig. 146

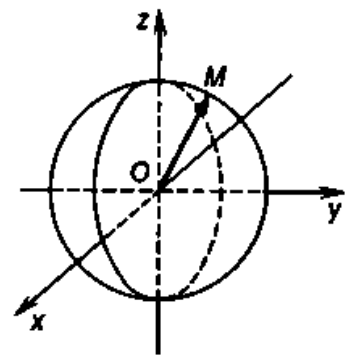



Fig. 147

*$M(x, y, z) \in S$ and is not satisfied by the coordinates of any point which does not lie on this surface.*

From the point of view of this definition, the surface $S$ is a set of points whose coordinates satisfy equation (1).

**Example.** In a rectangular system of coordinates the equation

$$x^2 + y^2 + z^2 - R^2 = 0,$$

or

$$x^2 + y^2 + z^2 = R^2$$

defines a surface which is a sphere of radius $R$ with centre at a point $O$ (0, 0, 0) (Fig. 147).

Indeed, if $M$ $(x, y, z)$ is an arbitrary point, then from formula (7) (see 2.9.2, Item 5)

$$|\overrightarrow{OM}| = \sqrt{x^2+y^2+z^2}.$$

Consequently, the given equation is satisfied by the coordinates of those and only those points which are at a distance $R$ from the point $O$. Thus the set of points whose coordinates satisfy this equation is a sphere of radius $R$ with centre at the origin.

We can consider a curve in space to be an intersection of two surfaces, i.e. a set of points which lie simultaneously on two surfaces, and correspondingly define the curve by two equations. *Thus two equations*

$$\begin{cases} F_1(x, y, z) = 0, \\ F_2(x, y, z) = 0 \end{cases} \tag{2}$$

*are the equations of the curve $L$ if they are satisfied by the coordinates of any point which lies on $L$ and are not satisfied by the coordinates of any point which does not lie on $L$.*

For example, the equations of two spheres

$$\begin{cases} x^2+y^2+(z-3)^2 = 10. \\ x^2+y^2+z^2 = 1 \end{cases}$$

simultaneously define a circle which lies in the $xy$-plane, whose radius is equal to unity and whose centre is at the origin.

### 2.9.10. The Equation of a Cylindrical Surface

Let a certain curve $L$ lie in the $xy$-plane (Fig. 148). Through every point of the curve $L$ we draw a straight line parallel to the $z$-axis. The set of these straight lines forms a surface $S$ which is said to be *cylindrical*. The indicated straight lines are the *generatrices* of the surface $S$ and the curve $L$ is its *directrix*.

We can define by analogy the cylindrical surface with generatrices which are parallel to the $x$-axis and $y$-axis.

For definiteness, we shall consider a cylindrical surface $S$ with generatrices parallel to the $z$-axis and prove that it is defined by an equation of the form

$$F(x, y) = 0. \tag{1}$$

Indeed, let (1) be an equation of the directrix $L$. We take on $S$ any point $M(x, y, z)$ which lies on some generatrix. If $M_0$ is the intersection of this generatrix and the $xy$-plane, then the point $M_0 \in L$ and its coordinates $x$ and $y$ satisfy equation (1). But then the numbers $x$, $y$, $z$ also satisfy this equation since $F(x, y)$ does not depend on $z$. Thus the coordinates $x$, $y$, $z$ of the arbitrary point

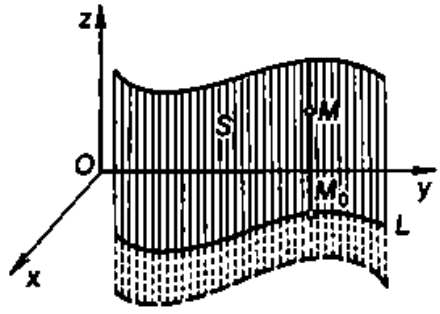



Fig. 148

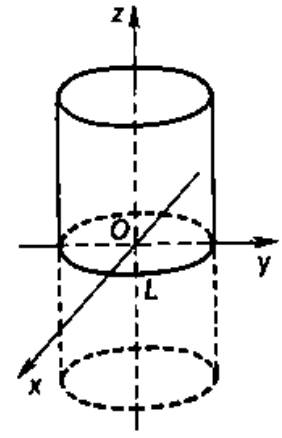



Fig. 149

$M \in S$ satisfy equation (1). It is evident that if $M(x, y, z) \notin S$, then $M_0(x, y) \notin L$, i.e. the coordinates $x$ and $y$ do not satisfy equation (1). This proves that (1) is an equation of the surface $S$.

Thus the equation of a cylindrical surface with generatrices parallel to the $z$-axis does not contain the coordinate $z$ and coincides with the equation of the directrix. For instance, if the ellipse

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1 \tag{2}$$

is the directrix, then the corresponding cylindrical surface is an *elliptic cylinder* and (2) is its equation.

Note that on the $xy$-plane the equation $F(x, y) = 0$ defines the curve $L$, but in the space system of coordinates $Oxyz$ the same curve $L$ is defined by two equations

$$\begin{cases} F(x, y)=0, \\ z=0. \end{cases}$$

Thus, for instance, in the space system of coordinates $Oxyz$ the equation $x^2 + y^2 - R^2 = 0$ defines a cylindrical surface which is a *circular cylinder* (Fig. 149), and the directrix $L$ of this cylinder (a circle), which lies in the $xy$-plane, is defined by two equations

$$\begin{cases} x^2+y^2=R^2, \\ z=0. \end{cases}$$

### 2.9.11. Equations of a Plane

In this chapter we show that first-order surfaces are planes and only planes and consider two kinds of equations of planes.

**1. The general equation of a plane.** Consider a rectangular system of coordinates $Oxyz$, an arbitrary plane $\pi$, a point $M_0(x_0, y_0, z_0) \in \pi$ and a vector $\mathbf{N} = \{A, B, C\}$ which is perpendicular to the plane $\pi$, where $A, B, C$ are its coordinates (Fig. 150).

We consider an arbitrary point $M(x, y, z)$. It lies on the plane $\pi$ if and only if the vectors $\overrightarrow{M_0M}$ and $\mathbf{N}$ are mutually perpendicular. Since the coordinates of the vector $\overrightarrow{M_0M}$ are equal to $x - x_0$, $y - y_0$, $z - z_0$, it follows, by virtue of the condition for the perpendicularity of two vectors (see formula (3) in 2.9.6), that the point $M(x, y, z)$ lies on the plane $\pi$ if and only if

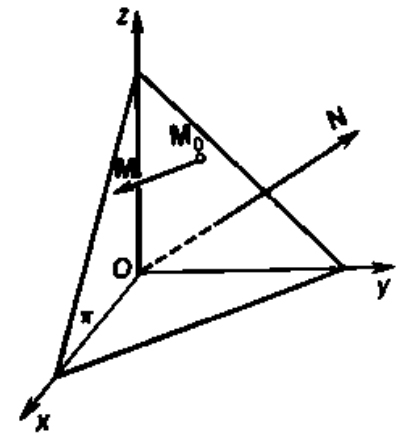


Fig. 150

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0. \tag{1}$$

And this is the required equation of the plane $\pi$ since it is satisfied by the coordinates $x$, $y$, $z$ of any point $M$ which lies on the plane $\pi$ and is not satisfied by the coordinates of any point which does not lie on this plane.

Removing the brackets, we reduce equation (1) to the form

$$Ax + By + Cz + (-Ax_0 - By_0 - Cz_0) = 0.$$

Next, designating the number $-Ax_0 - By_0 - Cz_0$ as $D$, we get

$$Ax + By + Cz + D = 0. \tag{2}$$

Equation (2) is known as the *general equation of a plane*. Thus a plane is a first-order surface since it is defined by a first-degree equation.

The converse is also true: every first-degree equation of form (2) defines a plane in a given rectangular system of coordinates.

Indeed, assume that we are given a rectangular system of coordinates $Oxyz$ and an equation $Ax + By + Cz + D = 0$ with arbitrary coefficients $A$, $B$, $C$, $D$, at least one of the coefficients $A$, $B$, $C$ being nonzero. This equation obviously has at least one solution $x_0$, $y_0$, $z_0$ $\left(\text{if, say, } C \neq 0\text{, we can take arbitrary } x_0 \text{ and } y_0 \text{ and get, from this equation, } z_0 = -\frac{A}{C}x_0 - \frac{B}{C}y_0 - \frac{D}{C}\right)$. Thus there is at least one point $M_0(x_0, y_0, z_0)$ whose coordinates satisfy the equation, i.e. $Ax_0 + By_0 + Cz_0 + D = 0$. Subtracting this

numerical equality from the equation $Ax + By + Cz + D = 0$, we get an equation

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0,$$

which is equivalent to the given equation. The equation obtained (and, hence, the equation $Ax + By + Cz + D = 0$) coincides with equation (1) and thus defines the plane $\pi$ which passes through the point $M_0(x_0, y_0, z_0)$ at right angles to the vector $\mathbf{N} = \{A, B, C\}$.

The vector $\mathbf{N} = \{A, B, C\}$ which is perpendicular to the plane is a *normal vector* of that plane.

**Example.** Set up an equation of the plane which passes through the point $M_0(1, 1, 1)$ at right angles to the vector $\mathbf{N} = \{2, 2, 3\}$.

*Solution.* From equation (1) we get the required equation

$$2(x - 1) + 2(y - 1) + 3(z - 1) = 0$$

or

$$2x + 2y + 3z - 7 = 0.$$

In conclusion, we shall prove the following theorem.

**Theorem 9.10.** *If two equations $A_1x + B_1y + C_1z + D_1 = 0$ and $A_2x + B_2y + C_2z + D_2 = 0$ define one and the same plane, then their coefficients are proportional.*

*Proof.* Indeed, the vectors $\mathbf{N}_1 = \{A_1, B_1, C_1\}$ and $\mathbf{N}_2 = \{A_2, B_2, C_2\}$ are perpendicular to this plane and, consequently, are collinear. But then the numbers $A_1, B_1, C_1$ are proportional to the numbers $A_2, B_2, C_2$ (see formula (2) in 2.9.4), i.e.

$$\frac{A_2}{A_1} = \frac{B_2}{B_1} = \frac{C_2}{C_1} = \mu,$$

or $A_2 = \mu A_1$, $B_2 = \mu B_1$, $C_2 = \mu C_1$ ($\mu$ is the proportionality factor). Multiplying the first of the given equations by $\mu$ and subtracting the result from the second equation, we get $D_2 - \mu D_1 = 0$, i.e. $D_2 = \mu D_1$ and, consequently,

$$\frac{A_2}{A_1} = \frac{B_2}{B_1} = \frac{C_2}{C_1} = \frac{D_2}{D_1} = \mu. \quad \blacksquare$$

**2. The angle between two planes.** Consider two planes $\pi_1$ and $\pi_2$ defined by the equations

$$A_1x + B_1y + C_1z + D_1 = 0, \quad A_2x + B_2y + C_2z + D_2 = 0$$

respectively.

For any mutual positions of the planes $\pi_1$ and $\pi_2$ in space, one of the angles $\varphi$ between them is equal to the angle between their normal vectors $\mathbf{N}_1 = \{A_1, B_1, C_1\}$ and $\mathbf{N}_2 = \{A_2, B_2, C_2\}$ and can be cal-

culated by the following formula:

$$\cos\varphi = \frac{\mathbf{N}_1 \cdot \mathbf{N}_2}{|\mathbf{N}_1|\,|\mathbf{N}_2|} = \frac{A_1A_2 + B_1B_2 + C_1C_2}{\sqrt{A_1^2+B_1^2+C_1^2}\,\sqrt{A_2^2+B_2^2+C_2^2}}. \tag{3}$$

The second angle is equal to $180° - \varphi$.

**3. The condition for parallelism of planes.** If the planes $\pi_1$ and $\pi_2$ are parallel, then their normal vectors $\mathbf{N}_1$ and $\mathbf{N}_2$ are collinear, and vice versa. But then

$$\frac{A_1}{A_2} = \frac{B_1}{B_2} = \frac{C_1}{C_2}. \tag{4}$$

Relation (4) is the condition for parallelism of the planes $\pi_1$ and $\pi_2$.

**4. The condition for perpendicularity of planes.** If the planes $\pi_1$ and $\pi_2$ are mutually perpendicular, then their normal vectors $\mathbf{N}_1$ and $\mathbf{N}_2$ are also mutually perpendicular ($\varphi = \pi/2$), and vice versa. Therefore formula (3) immediately yields the condition for the perpendicularity of the planes $\pi_1$ and $\pi_2$, i.e.

$$A_1A_2 + B_1B_2 + C_1C_2 = 0.$$

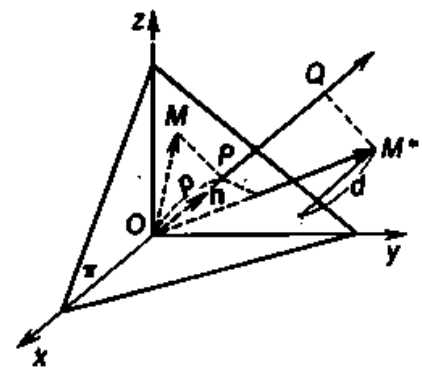


Fig. 151

**5. The normal equation of a plane. The distance from a point to a plane.** Assume that we are given a rectangular system of coordinates $Oxyz$ and an arbitrary plane $\pi$ (Fig. 151). Through the origin we draw a straight line at right angles to the plane $\pi$. We shall call it a *normal*. We designate the point at which the normal meets the plane $\pi$ as $P$. On the normal we introduce the direction from the point $O$ to the point $P$. If the points $O$ and $P$ coincide, then we take any one of the two directions on the normal. Assume that $\alpha$, $\beta$, $\gamma$ are the angles which the directed normal makes with the coordinate axes and $p$ is the length of the segment $OP$.

We shall derive an equation of the plane $\pi$ assuming the numbers $\cos\alpha$, $\cos\beta$, $\cos\gamma$, $p$ to be known. For that purpose, we introduce a unit vector $\mathbf{n}$ on the normal whose direction coincides with the positive direction of the normal.

Since $\mathbf{n}$ is a unit vector, it follows that

$$\mathbf{n} = \{\cos\alpha,\ \cos\beta,\ \cos\gamma\}. \tag{5}$$

Let $M(x, y, z)$ be an arbitrary point. It lies on the plane $\pi$ if and only if the projection of the vector $\overrightarrow{OM}$ on the normal is $p$, i.e.

$$\mathrm{proj}_{\mathbf{n}}\,\overrightarrow{OM} = p. \tag{6}$$

We note now that $\text{proj}_{\mathbf{n}} \overrightarrow{OM} = \mathbf{n} \cdot \overrightarrow{OM}$ and $\overrightarrow{OM} = \{x, y, z\}$. From Theorem 9.6, taking relation (5) into account, we have

$$\text{proj}_{\mathbf{n}} \overrightarrow{OM} = \mathbf{n} \cdot \overrightarrow{OM} = x \cos \alpha + y \cos \beta + z \cos \gamma. \tag{7}$$

From relations (6) and (7) we find that the point $M(x, y, z)$ lies on the plane $\pi$ if and only if its coordinates satisfy the equation

$$x \cos \alpha + y \cos \beta + z \cos \gamma - p = 0, \tag{8}$$

which is the required equation of the given plane. The equation of a plane in form (8) is called a *normal equation.*

**Theorem 9.11.** *If the point $M^*$ has coordinates $x^*$, $y^*$, $z^*$ and the plane is defined by the normal equation*

$$x \cos \alpha + y \cos \beta + z \cos \gamma - p = 0,$$

*then the distance $d$ from the point $M^*$ to that plane can be found from the formula*

$$d = | x^* \cos \alpha + y^* \cos \beta + z^* \cos \gamma - p |.$$

*Proof.* Let $Q$ be the projection of the point $M^*$ on the directed normal shown in Fig. 151. Then, by virtue of the fundamental identity (see 1.1.3), $PQ = OQ - OP$, whence it follows that

$$d = | PQ | = | OQ - OP |.$$

But $OQ = \text{proj}_{\mathbf{n}} \overrightarrow{OM^*}$, $OP = p$ and, consequently,

$$d = |\text{proj}_{\mathbf{n}} \overrightarrow{OM^*} - p|. \tag{9}$$

The vector $\overrightarrow{OM^*} = \{x^*, y^*, z^*\}$ and $\text{proj}_{\mathbf{n}} \overrightarrow{OM^*} = \mathbf{n} \cdot \overrightarrow{OM^*}$. According to Theorem 9.5, taking relation (5) into account, we find that

$$\text{proj}_{\mathbf{n}} \overrightarrow{OM^*} = \mathbf{n} \cdot \overrightarrow{OM^*} = x^* \cos \alpha + y^* \cos \beta + z^* \cos \gamma. \tag{10}$$

From relations (9) and (10) we finally get

$$d = | x^* \cos \alpha + y^* \cos \beta + z^* \cos \gamma - p |. \blacksquare$$

We shall show now how the general equation of a plane can be reduced to the normal form. Assume that

$$Ax + By + Cz + D = 0 \tag{11}$$

is the general equation of a plane, and

$$x \cos \alpha + y \cos \beta + z \cos \gamma - p = 0 \tag{12}$$

is its normal equation. Since equations (11) and (12) define the same plane, it follows, according to Theorem 9.10, that the coefficients

in these equations are proportional. This means that multiplying all the terms of (11) by a factor $\mu$, we get an equation

$$\mu Ax + \mu By + \mu Cz + \mu D = 0,$$

which coincides with equation (12), i.e. we have

$$\mu A = \cos\alpha,\ \mu B = \cos\beta,\ \mu C = \cos\gamma,\ \mu D = -p. \qquad (13)$$

To find the factor $\mu$, we square the first three equalities (13) and add the results together. Then we obtain

$$\mu^2 (A^2 + B^2 + C^2) = \cos^2\alpha + \cos^2\beta + \cos^2\gamma.$$

But in accordance with formula (6) of 2.9.2, the right-hand side of the last relation is equal to unity and, consequently,

$$\mu = \pm \frac{1}{\sqrt{A^2+B^2+C^2}}.$$

The number $\mu$ by means of which the general equation of a plane can be reduced to a normal equation, is known as a *normalizing factor* of that equation. The sign of $\mu$ is defined by the relation $\mu D = -p$, i.e. $\mu$ has the sign opposite to that of the constant term of the general equation (11).

If $D = 0$ in equation (11), then the sign of the normalizing factor can be arbitrarily chosen.

**Example.** Given a plane $x + 2y + 2z - 8 = 0$ and a point $M^*$ (1, 1, 1), find the distance $d$ from the point $M^*$ to the given plane.

*Solution.* To use Theorem 9.11, we must first reduce the given equation to the normal form. For that purpose, we find the normalizing factor

$$\mu = \frac{1}{\sqrt{1^2+2^2+2^2}} = \frac{1}{3}.$$

Multiplying the given equation by $\mu$, we get the required normal equation of the plane

$$\frac{1}{3}x + \frac{2}{3}y + \frac{2}{3}z - \frac{8}{3} = 0.$$

Substituting the coordinates of the point $M^*$ into the left-hand side of this equation, we obtain

$$d = \left|\frac{1}{3} + \frac{2}{3} + \frac{2}{3} - \frac{8}{3}\right| = \left|-\frac{3}{3}\right| = 1.$$

### 2.9.12. The Equation of a Straight Line

We have pointed out earlier that we can regard a curve in space as the intersection of two surfaces and define it by two equations. In particular, we can regard every straight line as an intersection of two surfaces and correspondingly define it by two first-degree equations.

Assume that we are given a rectangular system of coordinates $Oxyz$ and an arbitrary straight line $L$. We designate as $\pi_1$ and $\pi_2$ two different planes which intersect along the straight line $L$ and which are defined by the equations

$$A_1x + B_1y + C_1z + D_1 = 0, \quad A_2x + B_2y + C_2z + D_2 = 0 \quad (1)$$

respectively. Two equations of form (1) simultaneously define the line $L$ if and only if the planes $\pi_1$ and $\pi_2$ are nonparallel and noncoincident, i.e. the normal vectors $\mathbf{N}_1 = \{A_1, B_1, C_1\}$ and $\mathbf{N}_2 = \{A_2, B_2, C_2\}$ of these planes are noncollinear (the coefficients $A_1, B_1, C_1$ are not proportional to the coefficients $A_2, B_2, C_2$).

Equations (1) are known as the *general equations of a straight line.*

**1. Canonical equations of a straight line.** Equations (1) are not always convenient for problem solving and therefore a special form of the equation of a straight line is of use.

Consider a straight line $L$ and a nonzero vector $\mathbf{a}$ which lies on the given straight line or is parallel to it (Fig. 152). The vector $\mathbf{a}$ is a

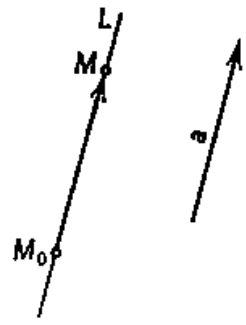



Fig. 152

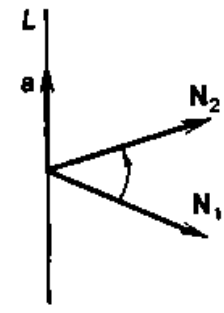



Fig. 153

*direction vector* of the straight line. We shall derive equations of the straight line which passes through the given point $M_0(x_0, y_0, z_0)$ and has the given direction vector $\mathbf{a} = \{l, m, n\}$ (see Fig. 152).

Let $M(x, y, z)$ be an arbitrary point. It lies on the straight line if and only if the vector $\overrightarrow{M_0M} = \{x - x_0, y - y_0, z - z_0\}$ is collinear with the direction vector $\mathbf{a} = \{l, m, n\}$, i.e. if the coordinates of the vector $\overrightarrow{M_0M}$ are proportional to the coordinates of the vector $\mathbf{a}$:

$$\frac{x - x_0}{l} = \frac{y - y_0}{m} = \frac{z - z_0}{n}. \quad (2)$$

Equations (2) are the required ones. They are the *canonical equations of a straight line.*

To derive the canonical equations (2) when the straight line $L$ is defined by equations (1), it is necessary

(1) to find a point $M_0(x_0, y_0, z_0) \in L$, for which purpose we must define the numerical value of one of the unknown coordinates, i.e.

$x_0$, $y_0$, $z_0$, and substitute it for the corresponding variable into equations (1) and then find the other two coordinates by simultaneously solving equations (1);

(2) to find the direction vector $\mathbf{a} = \{l, m, n\}$. Since the straight line $L$ is defined by the intersection of the planes $\pi_1$ and $\pi_2$, it is perpendicular to each of the normal vectors $\mathbf{N}_1$ and $\mathbf{N}_2$ (Fig. 153). Therefore we can take any vector perpendicular to the vectors $\mathbf{N}_1$ and $\mathbf{N}_2$, say, the vector product $\mathbf{a} = \mathbf{N}_1 \times \mathbf{N}_2$, as the vector $\mathbf{a}$. Since the coordinates of the vectors $\mathbf{N}_1$ and $\mathbf{N}_2$ are known, they are $\mathbf{N}_1 = \{A_1, B_1, C_1\}$ and $\mathbf{N}_2 = \{A_2, B_2, C_2\}$, it follows, according to Theorem 9.7, that the coordinates of the vector $\mathbf{a}$ are

$$\mathbf{a} = \left\{\begin{vmatrix} B_1 & C_1 \\ B_2 & C_2 \end{vmatrix}, \begin{vmatrix} C_1 & A_1 \\ C_2 & A_2 \end{vmatrix}, \begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix}\right\} = \{l, m, n\}.$$

**Example 1.** Find the canonical equations of the straight line

$$\begin{cases} 3x + 2y + 4z - 11 = 0, \\ 2x + y - 3z - 1 = 0. \end{cases}$$

*Solution.* Setting $x_0 = 1$, for instance, we find from the system

$$\begin{cases} 2y_0 + 4z_0 - 8 = 0, \\ y_0 - 3z_0 + 1 = 0 \end{cases}$$

that $y_0 = 2$ and $z_0 = 1$. Thus we have found the point $M_0$ (1, 2, 1) of the straight line. We shall now find the direction vector $\mathbf{a}$. We have $\mathbf{N}_1 = \{3, 2, 4\}$ and $\mathbf{N}_2 = \{2, 1, -3\}$, whence it follows that $\mathbf{a} = \mathbf{N}_1 \times \mathbf{N}_2 = \{-10, 17, -1\}$, i.e. $l = -10$, $m = 17$, $n = -1$. Substituting the values found for $x_0$, $y_0$, $z_0$ and $l$, $m$, $n$ into equations (2), we get the canonical equations of the given straight line:

$$\frac{x-1}{-10} = \frac{y-2}{17} = \frac{z-1}{-1}.$$

**2. Parametric equations of a straight line.** It is sometimes useful to represent a straight line in a form other than that of the canonical equations (2). Let the straight line $L$ be defined by equations (2). We designate each of the equal ratios as $t$. Then

$$\frac{x-x_0}{l} = \frac{y-y_0}{m} = \frac{z-z_0}{n} = t,$$

whence

$$x = x_0 + lt, \quad y = y_0 + mt, \quad z = z_0 + nt. \tag{3}$$

Equations (3) are the *parametric equations of the straight line* $L$ which passes through the point $M_0(x_0, y_0, z_0)$ and has the given direction vector $\mathbf{a} = \{l, m, n\}$. In equations (3) we regard $t$ as an arbitrarily varying parameter $(-\infty < t < +\infty)$ and $x$, $y$, $z$ as

functions of $t$. When $t$ varies, the quantities $x$, $y$, $z$ vary so that the point $M(x, y, z)$ moves along the straight line.

Parametric equations are convenient when we have to find the point of intersection of a straight line and a plane. Indeed, let the nonparallel plane $\pi$ and the straight line $L$ be defined by the relations

$$Ax + By + Cz + D = 0$$

and

$$x = x_0 + lt, \quad y = y_0 + mt, \quad z = z_0 + nt$$

respectively. To find the point of intersection of the straight line and the plane, we substitute the expressions for $x$, $y$, $z$ from the equations for $L$ into the equation for $\pi$. As a result of the transformations, we have

$$t = -\frac{Ax_0 + By_0 + Cz_0 + D}{Al + Bm + Cn},$$

the denominator of the fraction being nonzero since the plane is not parallel to the straight line (see 2.9.13). Substituting the value of $t$ obtained into the equations of the straight line, we find the required point $M(x, y, z)$ of the intersection of the line $L$ and the plane $\pi$.

**3. The angle between straight lines.** Consider two straight lines $L_1$ and $L_2$ defined by the equations

$$\frac{x-x_1}{l_1} = \frac{y-y_1}{m_1} = \frac{z-z_1}{n_1}, \quad \frac{x-x_2}{l_2} = \frac{y-y_2}{m_2} = \frac{z-z_2}{n_2}$$

respectively. For any position of the straight lines $L_1$ and $L_2$ in space, one of the two angles between them is equal to the angle $\varphi$ between their direction vectors $\mathbf{a}_1 = \{l_1, m_1, n_1\}$ and $\mathbf{a}_2 = \{l_2, m_2, n_2\}$ and the other angle is equal to $180° - \varphi$. The angle $\varphi$ can be found from the formula

$$\cos\varphi = \frac{l_1l_2 + m_1m_2 + n_1n_2}{\sqrt{l_1^2 + m_1^2 + n_1^2}\sqrt{l_2^2 + m_2^2 + n_2^2}}.$$

**4. The condition for parallelism of straight lines.** The straight lines $L_1$ and $L_2$ are parallel if and only if their direction vectors $\mathbf{a}_1 = \{l_1, m_1, n_1\}$ and $\mathbf{a}_2 = \{l_2, m_2, n_2\}$ are collinear. From this we get the condition for the parallelism of the straight lines $L_1$ and $L_2$:

$$\frac{l_2}{l_1} = \frac{m_2}{m_1} = \frac{n_2}{n_1}.$$

**5. The condition for perpendicularity of straight lines.** The straight lines $L_1$ and $L_2$ are perpendicular if and only if their direction vectors $\mathbf{a}_1 = \{l_1, m_1, n_1\}$ and $\mathbf{a}_2 = \{l_2, m_2, n_2\}$ are perpendicular. From this we get the condition for the perpendicularity of the straight lines $L_1$ and $L_2$:

$$l_1l_2 + m_1m_2 + n_1n_2 = 0.$$

**6. The distance from a point to a straight line.** In conclusion, we shall consider the problem of finding the distance $d$ from a given point to a given straight line in space.

Let us consider a straight line $L$: $\frac{x-x_0}{l}=\frac{y-y_0}{m}=\frac{z-z_0}{n}$ and a point $M_1(x_1, y_1, z_1)$. The required distance $d$ is the altitude of the parallelogram constructed on the vectors $\mathbf{a}=\{l, m, n\}$ and $\overrightarrow{M_0M_1}=\{x_1-x_0, y_1-y_0, z_1-z_0\}$ (Fig. 154). Let the vector $\mathbf{p}$ be the vector product of the vectors $\overrightarrow{M_0M_1}$ and $\mathbf{a}$, i.e. $\mathbf{p}=\overrightarrow{M_0M_1}\times\mathbf{a}$. Since $|\mathbf{p}|$ is equal to the area of the parallelogram constructed on the vectors $\overrightarrow{M_0M_1}$ and $\mathbf{a}$, we have

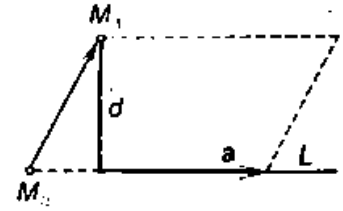


Fig. 154

$$d=\frac{|\mathbf{p}|}{|\mathbf{a}|}, \text{ where } |\mathbf{p}|$$

$$=\sqrt{\begin{vmatrix} y_1-y_0 & z_1-z_0 \\ m & n \end{vmatrix}^2+\begin{vmatrix} z_1-z_0 & x_1-x_0 \\ n & l \end{vmatrix}^2+\begin{vmatrix} x_1-x_0 & y_1-y_0 \\ l & m \end{vmatrix}^2},$$

$$|\mathbf{a}|=\sqrt{l^2+m^2+n^2}.$$

Consequently,

$$d=\frac{\sqrt{\begin{vmatrix} y_1-y_0 & z_1-z_0 \\ m & n \end{vmatrix}^2+\begin{vmatrix} z_1-z_0 & x_1-x_0 \\ n & l \end{vmatrix}^2+\begin{vmatrix} x_1-x_0 & y_1-y_0 \\ l & m \end{vmatrix}^2}}{\sqrt{l^2+m^2+n^2}}.$$

### 2.9.13. Mutual Positions of a Straight Line and a Plane

**1. The conditions for parallelism and perpendicularity.** Consider a straight line

$$\frac{x-x_0}{l}=\frac{y-y_0}{m}=\frac{z-z_0}{n}$$

and a plane

$$Ax+By+Cz+D=0.$$

The straight line is parallel to the plane if and only if its direction vector $\mathbf{a}=\{l, m, n\}$ is perpendicular to the normal vector $\mathbf{N}=\{A, B, C\}$ of the plane. From this we get the condition for parallelism of a straight line and a plane:

$$Al+Bm+Cn=0.$$

The straight line is perpendicular to the plane if and only if its direction vector is collinear with the normal vector of the plane. From this we get the condition for perpendicularity of a straight

line and a plane:

$$\frac{A}{l}=\frac{B}{m}=\frac{C}{n}.$$

**2. The angle between a straight line and a plane.** Let us consider a plane $\pi$

$$Ax+By+Cz+D=0$$

and a straight line $L$

$$\frac{x-x_0}{l}=\frac{y-y_0}{m}=\frac{z-z_0}{n}$$

which is not perpendicular to the plane. The angle $\varphi$ between the straight line $L$ and the plane $\pi$ is the acute angle between $L$ and its projection on $\pi$ (Fig. 155). We designate the angle between the vectors $\mathbf{N}=\{A, B, C\}$ and $\mathbf{a}=\{l, m, n\}$ as $\theta$. If $\theta \leqslant 90°$ (as in Fig. 155), then $\varphi = 90° - \theta$ and $\sin\varphi = \sin(90° - \theta) = \cos\theta$. Now if $\theta > 90°$, then $\varphi = \theta - 90°$ and $\sin\varphi = \sin(\theta - 90°) = -\cos\theta$. In any case $\sin\varphi = |\cos\theta|$. But the formula for $\cos\theta$ is known (see formula (4) in 2.9.6) and, consequently,

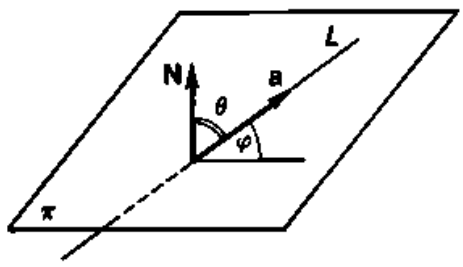



Fig. 155

$$\sin\varphi=\frac{|Al+Bm+Cn|}{\sqrt{A^2+B^2+C^2}\sqrt{l^2+m^2+n^2}}.$$

### 2.9.14. Second-Order Surfaces

Second-order surfaces are surfaces which, in a rectangular system of coordinates, are defined by second-degree algebraic equations.

We shall employ the method of parallel sections to carry out a geometric investigation of second-order surfaces, using some given equations.

**1. Ellipsoid.** *An ellipsoid is a surface which, in a certain rectangular system of coordinates, is defined by the equation*

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1. \tag{1}$$

Equation (1) is the *canonical equation of an ellipsoid.*

We shall establish the shape of an ellipsoid. For that purpose, we consider its sections by planes parallel to the $xy$-plane. Each of these planes is defined by an equation of the form $z = h$, where $h$ is any number, and the curve resulting in the section is defined by two equations

$$\begin{cases}\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=1-\dfrac{h^2}{c^2},\\ z=h\end{cases} \tag{2}$$

from which it follows that

for $|h| > c$ $(c > 0)$, $\frac{x^2}{a^2} + \frac{y^2}{b^2} < 0$ and equations (2) define an imaginary ellipse, i.e. there are no points of intersection of the plane $z = h$ and the given ellipsoid;

for $h = \pm c$, $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 0$ and curve (2) degenerates into points $(0, 0, +c)$ and $(0, 0, -c)$ (the planes $z = \pm c$ touch the ellipsoid);

for $|h| < c$ equations (2) can be represented as

$$\begin{cases} \frac{x^2}{a^{*2}} + \frac{y^2}{b^{*2}} = 1, \\ z = h, \end{cases}$$

whence it follows that the plane $z = h$ cuts the ellipsoid along an ellipse with semi-axes $a^* = a\sqrt{1 - h^2/c^2}$ and $b^* = b\sqrt{1 - h^2/c^2}$. When $|h|$ diminishes, the values of $a^*$ and $b^*$ increase and attain their greatest values for $h=0$, i.e. when the ellipsoid is cut by the $xy$-plane, the section is the greatest ellipse with semi-axes $a^* = a$ and $b^* = b$.

A similar picture results when a surface is intersected by planes parallel to the $xz$-plane and $yz$-plane.

Thus the sections we have considered make it possible to represent an ellipsoid as a closed oval surface (Fig. 156).

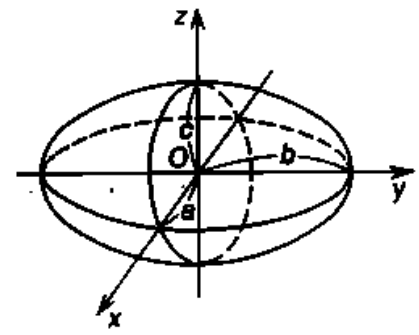


Fig. 156

The quantities $a$, $b$, $c$ are the *semi-axes* of the ellipsoid. In the case $a = b = c$ this ellipsoid is a *sphere*.

**2. Hyperboloid of one sheet.** *A hyperboloid of one sheet is a surface which, in a certain rectangular system of coordinates, is defined by the equation*

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1. \tag{3}$$

Equation (3) is the *canonical equation of a one-sheet hyperboloid.*

We shall find the shape of surface (3). For that purpose, we consider its sections by the $xz$-plane $(y = 0)$ and the $yz$-plane $(x = 0)$. Respectively, we get the following equations:

$$\begin{cases} \frac{x^2}{a^2} - \frac{z^2}{c^2} = 1, \\ y = 0 \end{cases} \quad \text{and} \quad \begin{cases} \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1, \\ x = 0 \end{cases}$$

from which it follows that the sections are hyperbolas.

We shall now consider the section of the given hyperboloid by the planes $z = h$ parallel to the $xy$-plane. The curve resulting in the

section is defined by the equations

$$\begin{cases} \frac{x^2}{a^2}+\frac{y^2}{b^2}=1+\frac{h^2}{c^2}, \\ z=h \end{cases} \quad \text{or} \quad \begin{cases} \frac{x^2}{a^{*2}}+\frac{y^2}{b^{*2}}=1, \\ z=h \end{cases} \tag{4}$$

from which it follows that the plane $z = h$ cuts the hyperboloid along an ellipse with semi-axes $a^* = a\sqrt{1 + h^2/c^2}$ and $b^* = b\sqrt{1 + h^2/c^2}$ which attain their least values for $h = 0$, i.e. the section of the given hyperboloid by the $xy$-plane is the smallest ellipse with

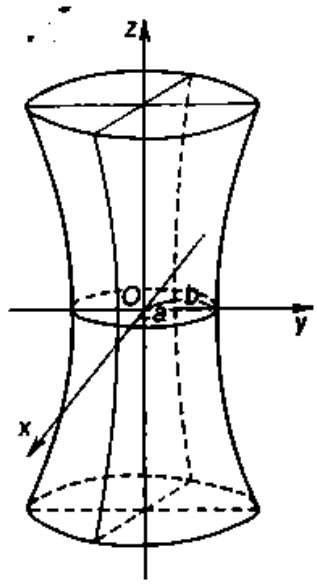



Fig. 157

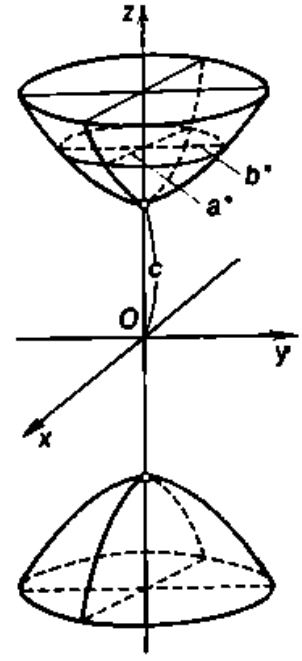



Fig. 158

semi-axes $a^* = a$ and $b^* = b$. When $|h|$ increases indefinitely, the quantities $a^*$ and $b^*$ increase indefinitely.

Thus the sections we have considered make it possible to represent a one-sheet hyperboloid as an infinite tube which becomes infinitely wider as it recedes (on both sides) from the $xy$-plane (Fig. 157).

The quantities $a$, $b$, $c$ are the *semi-axes* of the one-sheet hyperboloid. The first two of them are shown in Fig. 157, and to plot the semi-axis $c$ on the drawing, we must construct the basic rectangle of any one of the hyperbolas.

**3. Hyperboloid of two sheets.** *A hyperboloid of two sheets is a surface which, in a certain rectangular system of coordinates, is defined by the equation*

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=-1. \tag{5}$$

Equation (5) is the *canonical equation of a two-sheet hyperboloid.*

We shall establish the shape of surface (5), for which purpose we shall consider its sections by the $xz$-plane and the $yz$-plane. Respectively, we get equations

$$\begin{cases}\frac{x^2}{a^2}-\frac{z^2}{c^2}=-1,\\ y=0\end{cases} \quad \text{and} \quad \begin{cases}\frac{y^2}{b^2}-\frac{z^2}{c^2}=-1,\\ x=0\end{cases}$$

from which it follows that the sections are hyperbolas.

We shall now consider the sections of the given hyperboloid by the planes $z = h$ parallel to the $xy$-plane. The curve resulting in the section is defined by the equations

$$\begin{cases}\frac{x^2}{a^2}+\frac{y^2}{b^2}=\frac{h^2}{c^2}-1,\\ z=h\end{cases} \quad \text{or} \quad \begin{cases}\frac{x^2}{a^{*2}}+\frac{y^2}{b^{*2}}=1,\\ z=h\end{cases} \tag{6}$$

from which it follows that

for $|h| > c$ $(c > 0)$, the plane $z = h$ cuts the hyperboloid along an ellipse with semi-axes $a^* = a\sqrt{h^2/c^2-1}$ and $b^* = b\sqrt{h^2/c^2-1}$; as $|h|$ increases, the quantities $a^*$ and $b^*$ also increase;

for $h = \pm c$ equations (6) are satisfied by the coordinates of only two points, $(0, 0, +c)$ and $(0, 0, -c)$ (the planes $z = \pm c$ touch the given surface);

for $|h| < c$ equations (6) define an imaginary ellipse, i.e. there are no points of intersection of the plane $z = h$ and the hyperboloid.

Thus the sections we have considered make it possible to represent a two-sheet hyperboloid as a surface consisting of two separate "sheets" (hence its name), each of which is shaped as an infinite convex bowl (Fig. 158).

The quantities $a$, $b$, $c$ are the *semi-axes* of the two-sheet hyperboloid. Figure 158 shows the quantity $c$. To show $a$ and $b$ on the drawing, we must construct the basic rectangles of the hyperbolas in the $xz$-plane and $yz$-plane.

**4. Elliptic paraboloid.** *An elliptic paraboloid is a surface which, in a certain rectangular system of coordinates, is defined by the equation*

$$\frac{x^2}{p}+\frac{y^2}{q}=2z, \tag{7}$$

*where* $p > 0$ *and* $q > 0$.

Equation (7) is the *canonical equation of an elliptic paraboloid.*

We shall use various sections to investigate this surface. We shall first consider the sections of the given surface by the $xz$-plane and the $yz$-plane. Respectively, we get equations

$$\begin{cases}x^2=2pz,\\ y=0\end{cases} \quad \text{and} \quad \begin{cases}y^2=2qz,\\ x=0\end{cases}$$

from which it follows that the sections are parabolas which are symmetric about the $z$-axis and have vertices at the origin.

We shall now consider the sections of the given paraboloid by the planes $z = h$ which are parallel to the $xy$-plane. The curve resulting in the section is defined by the equations

$$\begin{cases} \frac{x^2}{p} + \frac{y^2}{q} = 2h, \\ z = h \end{cases} \quad \text{or} \quad \begin{cases} \frac{x^2}{a^{*2}} + \frac{y^2}{b^{*2}} = 1, \\ z = h \end{cases} \tag{8}$$

from which it follows that

for $h > 0$ the plane $z = h$ cuts the elliptic paraboloid along an ellipse with semi-axes $a^* = \sqrt{2hp}$ and $b^* = \sqrt{2hq}$; as $h$ increases, the quantities $a^*$ and $b^*$ also increase;

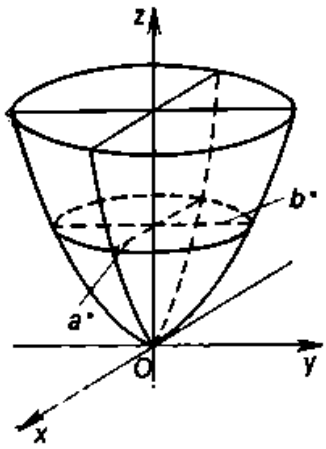


Fig. 159

for $h = 0$ the ellipse degenerates into a point (the plane $z = 0$ touches the paraboloid);

for $h < 0$ equations (8) define an imaginary ellipse, i.e. there are no points of intersection of the plane $z = h$ and the paraboloid.

Thus the sections we have considered make it possible to represent an elliptic paraboloid as an infinite convex bowl (Fig. 159).

The point (0, 0, 0) is the *vertex* of the elliptic paraboloid, and the numbers $p$ and $q$ are its *parameters*.

In the case $p = q$ equations (8) define a circle with centre on the $z$-axis, i.e. an elliptic paraboloid can be regarded as a surface formed by the rotation of a parabola about its axis. Such a surface is known as a *paraboloid of revolution*.

**5. Hyperbolic paraboloid.** *A hyperbolic paraboloid is a surface which, in a certain rectangular system of coordinates, is defined by the equation*

$$\frac{x^2}{p} - \frac{y^2}{q} = 2z, \tag{9}$$

*where* $p > 0$ *and* $q > 0$.

Equation (9) is the *canonical equation of a hyperbolic paraboloid.*

We shall establish the shape of surface (9), for which purpose we shall consider the section of the paraboloid by the $xz$-plane ($y = 0$). We get equations

$$\begin{cases} x^2 = 2pz, \\ y = 0 \end{cases} \tag{10}$$

from which it follows that the section is a parabola convex downwards, which is symmetric about the $z$-axis, with the vertex at the origin.

The sections of the surface by the planes $y = h$ parallel to the $xz$-plane are also parabolas convex downwards

$$\begin{cases} x^2 = 2p\left(z + \frac{h^2}{2q}\right), \\ y = h. \end{cases}$$

We shall consider the section of the paraboloid by the $yz$-plane ($x = 0$). We get equations

$$\begin{cases} y^2 = -2qz, \\ x = 0 \end{cases}$$

from which it follows that in this case as well the section is a parabola, but this time convex upwards, which is symmetric about the $z$-axis, with the vertex at the origin.

We shall consider the sections of the paraboloid by the planes $x = h$ parallel to the $yz$-plane and get equations

$$\begin{cases} y^2 = -2q\left(z - \frac{h^2}{2p}\right), \\ x = h \end{cases}$$

from which it follows that for any $h$ the section is a parabola convex upwards, its vertex lying on the parabola, defined by equations (10).

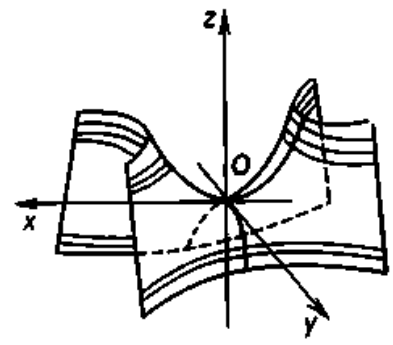


Fig. 160

We shall finally consider the sections of the paraboloid by the planes $z = h$ parallel to the $xy$-plane. We get equations

$$\begin{cases} \frac{x^2}{p} - \frac{y^2}{q} = 2h, \\ z = h \end{cases} \quad \text{or} \quad \begin{cases} \frac{x^2}{2ph} - \frac{y^2}{2qh} = 1, \\ z = h \end{cases}$$

from which it follows that

for $h > 0$ the sections are hyperbolas which cut the $xz$-plane;

for $h < 0$ they are hyperbolas which cut the $yz$-plane;

for $h = 0$ the hyperbola degenerates into a pair of intersecting straight lines

$$\begin{cases} \frac{x}{\sqrt{p}} - \frac{y}{\sqrt{q}} = 0, \\ z = 0 \end{cases} \quad \text{and} \quad \begin{cases} \frac{x}{\sqrt{p}} + \frac{y}{\sqrt{q}} = 0, \\ z = 0. \end{cases}$$

Thus the sections we have considered make it possible to represent a hyperbolic paraboloid as a saddle-like surface (Fig. 160). The figure shows several sections of the paraboloid by the planes $z = h$ for the cases $h > 0$ and $h < 0$.

The point (0, 0, 0) is the *vertex* of the hyperbolic paraboloid, and the numbers $p$ and $q$ are its *parameters*.

**6. Paraboloid cone.** *A second-order, or paraboloid, cone is a surface which, in a certain rectangular system of coordinates, is defined by the equation*

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=0. \tag{11}$$

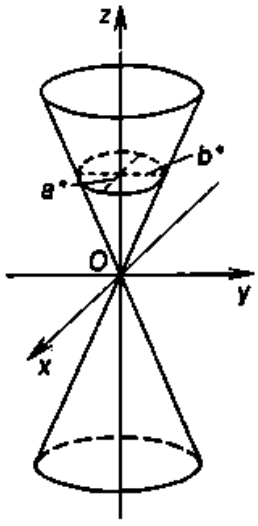


Fig. 161

Equation (11) is the *canonical equation of a paraboloid cone*.

We shall consider the geometric properties of a cone. In the section of this surface by the $xz$-plane ($y = 0$) we get a curve

$$\begin{cases}\frac{x^2}{a^2}-\frac{z^2}{c^2}=0,\\ y=0\end{cases}$$

which breaks up into two intersecting straight lines

$$\begin{cases}\frac{x}{a}+\frac{z}{c}=0,\\ y=0\end{cases} \quad \text{and} \quad \begin{cases}\frac{x}{a}-\frac{z}{c}=0,\\ y=0.\end{cases}$$

Similarly, the section of a cone by the $yz$-plane ($x = 0$) is also represented by two intersecting straight lines

$$\begin{cases}\frac{y}{b}+\frac{z}{c}=0,\\ x=0\end{cases} \quad \text{and} \quad \begin{cases}\frac{y}{b}-\frac{z}{c}=0,\\ x=0.\end{cases}$$

We shall consider now the sections of the given surface by the planes $z = h$ parallel to the $xy$-plane. We get equations

$$\begin{cases}\frac{x^2}{a^2}+\frac{y^2}{b^2}=\frac{h^2}{c^2},\\ z=h\end{cases} \quad \text{or} \quad \begin{cases}\frac{x^2}{a^{*2}}+\frac{y^2}{b^{*2}}=1,\\ z=h\end{cases}$$

from which it follows that

for $h > 0$ and $h < 0$ the sections are ellipses with semi-axes $a^* = \frac{a|h|}{c}$ and $b^* = \frac{b|h|}{c}$; with an increase in the absolute value of $h$ the semi-axes $a^*$ and $b^*$ also increase;

for $h = 0$ the line of intersection of the surface and the plane $z = h$ degenerates into a point (0, 0, 0).

Thus the sections we have considered make it possible to represent a cone as a surface shown in Fig. 161.

Chapter 10

# ELEMENTS OF HIGHER ALGEBRA

## 2.10.1. Matrices

**1. Definition of a matrix.** Assume that we are given a rectangular system of coordinates $Oxyz$ and a point $M$ with coordinates $x_1$, $x_2$, $x_3$*. We consider a radius vector** $\mathbf{r}$ of the point $M$: $\mathbf{r} = \overrightarrow{OM}$. We designate as $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ the unit base vectors, and then, in the given system of coordinates, the vector $\mathbf{r}$ is written as

$$\mathbf{r} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3.$$

We transform this system of coordinates rotating it about the origin $O$ and assume the angles which each axis of the new system of coordinates $Ox'y'z'$ makes with each axis of the old system to be known. We designate the coordinates of the point $M$ in the new coordinate system as $x_1'$, $x_2'$, $x_3'$, and the unit base vectors of the new axes as $\mathbf{e}_1'$, $\mathbf{e}_2'$, $\mathbf{e}_3'$. Then, in the new coordinate system, the vector $\mathbf{r}$ is written as

$$\mathbf{r} = x_1'\mathbf{e}_1' + x_2'\mathbf{e}_2' + x_3'\mathbf{e}_3'.$$

Equating the expressions for the vector $\mathbf{r}$ to one another, we get a vector equality

$$x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3 = x_1'\mathbf{e}_1' + x_2'\mathbf{e}_2' + x_3'\mathbf{e}_3'. \tag{1}$$

Let us see how the coordinates of the point $M$ are transformed when the coordinate system is rotated.

We resolve the vectors $\mathbf{e}_1'$, $\mathbf{e}_2'$, $\mathbf{e}_3'$ into components with respect to the old basis:

$$\begin{cases} \mathbf{e}_1' = a_{11}\mathbf{e}_1 + a_{21}\mathbf{e}_2 + a_{31}\mathbf{e}_3, \\ \mathbf{e}_2' = a_{12}\mathbf{e}_1 + a_{22}\mathbf{e}_2 + a_{32}\mathbf{e}_3, \\ \mathbf{e}_3' = a_{13}\mathbf{e}_1 + a_{23}\mathbf{e}_2 + a_{33}\mathbf{e}_3. \end{cases} \tag{2}$$

Each of the vectors $\mathbf{e}_1'$, $\mathbf{e}_2'$, $\mathbf{e}_3'$ is a unit vector and therefore for each of them the direction cosines are the coefficients of the expansion, i.e.

$$a_{11} = \cos(\widehat{\mathbf{e}_1, \mathbf{e}_1'}), \quad a_{12} = \cos(\widehat{\mathbf{e}_1, \mathbf{e}_2'}), \quad a_{13} = \cos(\widehat{\mathbf{e}_1, \mathbf{e}_3'}),$$

$$a_{21} = \cos(\widehat{\mathbf{e}_2, \mathbf{e}_1'}), \quad a_{22} = \cos(\widehat{\mathbf{e}_2, \mathbf{e}_2'}), \quad a_{23} = \cos(\widehat{\mathbf{e}_2, \mathbf{e}_3'}),$$

$$a_{31} = \cos(\widehat{\mathbf{e}_3, \mathbf{e}_1'}), \quad a_{32} = \cos(\widehat{\mathbf{e}_3, \mathbf{e}_2'}), \quad a_{33} = \cos(\widehat{\mathbf{e}_3, \mathbf{e}_3'}).$$

* In this section we designate the coordinates of the point $M$ as $x_1$, $x_2$, $x_3$ (rather than $x$, $y$, $z$).

** The radius vector of the point $M$ is a vector from the origin of coordinates to this point.

Replacing the vectors $\mathbf{e}_1'$, $\mathbf{e}_2'$, $\mathbf{e}_3'$ in relation (1) by their expansions (2) and grouping terms, we obtain

$$x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3 = (a_{11}x_1' + a_{12}x_2' + a_{13}x_3')\,\mathbf{e}_1$$
$$+ (a_{21}x_1' + a_{22}x_2' + a_{23}x_3')\,\mathbf{e}_2 + (a_{31}x_1' + a_{32}x_2' + a_{33}x_3')\mathbf{e}_3.$$

From this, equating the coefficients in the same base vectors to one another, we find that

$$\begin{cases} x_1 = a_{11}x_1' + a_{12}x_2' + a_{13}x_3', \\ x_2 = a_{21}x_1' + a_{22}x_2' + a_{23}x_3', \\ x_3 = a_{31}x_1' + a_{32}x_2' + a_{33}x_3'. \end{cases} \tag{3}$$

Consequently, the coordinates $x_1$, $x_2$, $x_3$ are linear combinations of the coordinates $x_1'$, $x_2'$, $x_3'$ which are completely defined by the collection of coefficients $a_{11}, \ldots, a_{33}$. We have thus arrived at the concept of a matrix.

**Definition.** *A table composed of coefficients* (3) *and written in the form*

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \tag{4}$$

*is a matrix of the given transformation.* In a brief form, the matrix can be written as

$$A = (a_{ij}) \quad (i = 1,\ 2,\ 3,\ j = 1,\ 2,\ 3),$$

where $a_{ij}$ are *elements* of this matrix.

Elements of a matrix form *rows* and *columns*. The first index ($i$) shows the number of the row and the second index ($j$) shows the number of the column whose intersection is occupied by the element $a_{ij}$. Matrix (4) has three rows and three columns.

Higher algebra studies matrices with any number of rows and columns and therefore in the general form a matrix is written as

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}. \tag{5}$$

If the number of rows in a matrix is equal to the number of columns ($m = n$), then it is a *square nth-order matrix*, otherwise it is a *rectangular matrix*. Thus matrix (4) is square, of order, or dimension, three. In matrix (5) there are $m$ rows and $n$ columns. If $m = 1$ and $n > 1$, then we get a single-row matrix

$$(a_1 a_2 \ldots a_n)$$

which is known as a *row vector*, or *row matrix*, and if $m > 1$ and $n = 1$, we get a single-column matrix

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}$$

which is known as a *column vector*, or *column matrix*.

Two matrices $A = (a_{ij})$ and $B = (b_{ij})$ are equal if the elements occupying the same places are equal, i.e. if $a_{ij} = b_{ij}$ for all $i$ and $j$ (in this case the number of rows (and, similarly, columns) of the matrices $A$ and $B$ must be equal).

**2. Properties of matrices.** Matrices, like vectors, can be added, multiplied by a number and one by another. Let us consider these operations.

1°. *The sum of two matrices $A = (a_{ij})$ and $B = (b_{ij})$ with the same number $m$ of rows and $n$ of columns is a matrix $C = (c_{ij})$ whose elements are defined by the relation*

$$a_{ij} + b_{ij} = c_{ij} \quad (i = 1, 2, \ldots, m; \quad j = 1, 2, \ldots, n).$$

The notation is $A + B = C$.

**Example 1.**

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1+0 & 0+0 & 0+0 \\ 0+0 & 1+0 & 0+0 \\ 0+0 & 0+0 & 0+1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The difference of two matrices can be defined by analogy.

2°. *The product of the matrix $A = (a_{ij})$ by the number $\lambda$ is a matrix whose every element is equal to the product of the respective element of the matrix $A$ by the number $\lambda$, i.e.*

$$\lambda A = \lambda (a_{ij}) = (\lambda a_{ij}) \quad (i = 1, 2, \ldots, m; \; j = 1, 2, \ldots, n).$$

**Example 2.**

$$3\begin{pmatrix} 1 & 0 & 2 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 3\cdot 1 & 3\cdot 0 & 3\cdot 2 \\ 3\cdot 2 & 3\cdot 0 & 3\cdot 0 \\ 3\cdot 0 & 3\cdot 1 & 3\cdot 0 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 6 \\ 6 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix}.$$

3°. *The product of the matrix $A = (a_{ij})$, which has $m$ rows and $k$ columns by the matrix $B = (b_{ij})$ which has $k$ rows and $n$ columns is a matrix $C = (c_{ij})$ which has $m$ rows and $n$ columns and whose element $c_{ij}$ is equal to the sum of the products of the elements of the $i$th row of the*

*matrix $A$ by the $j$th column of the matrix $B$, i.e.*

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{ik}b_{kj}$$
$$(i = 1, 2, \ldots, m; \quad j = 1, 2, \ldots, n).$$

In this case the number $k$ of columns of the matrix $A$ must be equal to the number of rows of the matrix $B$, otherwise the product is not defined. The product is designated as $A \cdot B = C$.

**Example 3.**

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 0\cdot 1 + 1\cdot 0 + 0\cdot 1 & 0\cdot 0 + 1\cdot 1 + 0\cdot 0 & 0\cdot 1 + 1\cdot 0 + 0\cdot 1 \\ 1\cdot 1 + 0\cdot 0 + 1\cdot 1 & 1\cdot 0 + 0\cdot 1 + 1\cdot 0 & 1\cdot 1 + 0\cdot 0 + 1\cdot 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 2 & 0 & 2 \end{pmatrix}.$$

**Example 4.** Let $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, then

$$A \cdot B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \text{ and } B \cdot A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

From this we find that

$$AB \neq BA,$$

i.e. the multiplication of matrices is not commutative.

**Remark.** It is easy to memorize the rule of multiplication by formulating it as follows: the element $c_{ij}$ of the matrix $C$ which is in the $i$th row and the $j$th column is the scalar product of the $i$th row vector of the matrix $A$ by the $j$th column vector of the matrix $B$.

It is easy to make sure by a direct verification that the following relations hold true for the sum and the product of matrices:

$$(A + B) \cdot C = A \cdot C + B \cdot C,$$
$$C \cdot (A + B) = C \cdot A + C \cdot B,$$
$$A \cdot (B \cdot C) = (A \cdot B) \cdot C,$$
$$(A + B) + C = A + (B + C).$$

4°. *Multiplication by an identity matrix.* The collection of elements $a_{11}, a_{22}, \ldots, a_{nn}$ of the square matrix $A = (a_{ij})$ is the *principal diagonal* of the matrix. A matrix whose elements occupying the principal diagonal are equal to unity and all other elements are zero is an *identity* (or *unit*) *matrix* and designated as $I$.

Thus the matrix

$$I=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix} \tag{6}$$

is a third-order identity matrix.

An identity matrix has a remarkable property, namely, *the multiplication of a square matrix of any order by the corresponding identity matrix does not change the matrix*. This property explains the name "unit" matrix, i.e. in the multiplication of matrices it possesses the same property as the number 1 in the multiplication of numbers.

**Example 5.** Let $A=\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}$ and $I=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$; then, in accordance with the rule of multiplication of matrices, we have

$$A\cdot I=\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}\cdot\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}=\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}$$

and

$$I\cdot A=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\cdot\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}=\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix},$$

whence

$$A\cdot I=A \quad \text{and} \quad I\cdot A=A.$$

Note that the identity matrix (6) is associated with the following transformation of the coordinates of a point:

$$x_1=x_1', \quad x_2=x_2', \quad x_3=x_3'.$$

This is an *identical transformation.*

The concept of a matrix is closely associated with the concept of a determinant.

### 2.10.2. Determinants

**1. Definition of a determinant.** Consider a third-order square matrix whose elements, for the sake of convenience, are designated as $a_1$, $a_2$, $a_3$, $b_1$, $b_2$, $b_3$, $c_1$, $c_2$, $c_3$:

$$\begin{pmatrix}a_1 & b_1 & c_1\\ a_2 & b_2 & c_2\\ a_3 & b_3 & c_3\end{pmatrix}. \tag{1}$$

**Definition.** *A third-order determinant corresponding to matrix* (1) *is a number designated as*

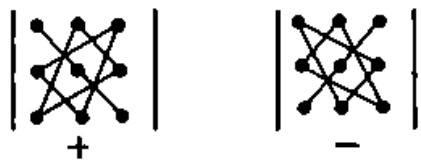
$$\Delta = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

*and defined by the relation*

$$\Delta = a_1b_2c_3 + b_1c_2a_3 + c_1a_2b_3 - c_1b_2a_3 - b_1a_2c_3 - a_1c_2b_3. \tag{2}$$

The numbers $a_1$, $a_2$, $a_3$, $b_1$, $b_2$, $b_3$, $c_1$, $c_2$, $c_3$ are the *elements* of the determinant. The diagonal composed of the elements $a_1$, $b_2$, $c_3$ is the *principal*, or *leading*, *diagonal* and the diagonal composed of the elements $a_3$, $b_2$, $c_1$ is the *secondary diagonal.*

To remember which products on the right-hand side of formula (2) must be taken with the plus sign and which with the minus sign, it is useful to employ the following rule of triangles:




This rule makes it easy to write formula (2) and calculate the determinant. For example,

$$\begin{vmatrix} 3 & -2 & 1 \\ -2 & 1 & 3 \\ 2 & 0 & -2 \end{vmatrix} = 3\cdot 1\cdot(-2) + (-2)\cdot 3\cdot 2 + (-2)\cdot 0\cdot 1$$
$$-2\cdot 1\cdot 1 - 3\cdot 0\cdot 3 - (-2)\cdot(-2)\cdot(-2) = -12.$$

**2. Properties of determinants.** We shall formulate and prove the properties for third-order determinants, although they are true for determinants of any order.

1°. *The value of a determinant does not vary if its rows and columns are interchanged, i.e.*

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}.$$

To prove this property, it is sufficient to apply formula (2) to the determinants on the left-hand and right-hand sides of the relation and make sure that the expressions obtained are equal.

Property 1° establishes the equivalence of the rows and columns of a determinant and therefore all other properties will be formulated

both for rows and for columns, but will be proved either for rows or for columns.

2°. *Interchanging of two rows or of two columns of a determinant is equivalent to its multiplication by* $-1$. For example,

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = - \begin{vmatrix} c_1 & b_1 & a_1 \\ c_2 & b_2 & a_2 \\ c_3 & b_3 & a_3 \end{vmatrix}.$$

This property can be proved by analogy with the preceding one.

3°. *If a determinant has two identical columns or two identical rows, then it is equal to zero.*

Indeed, when two identical columns are interchanged, the determinant $\Delta$ is not altered, but, according to Property 2, its sign changes to the opposite. Consequently, $\Delta = -\Delta$, i.e. $2\Delta = 0$, or $\Delta = 0$. For example,

$$\begin{vmatrix} 1 & 1 & 17 \\ 2 & 2 & 15 \\ 3 & 3 & 9 \end{vmatrix} = 0.$$

4°. *The multiplication of all elements of one column or of one row of a determinant by any number $\lambda$ is equivalent to the multiplication of the determinant by this number.* For example,

$$\begin{vmatrix} \lambda a_1 & b_1 & c_1 \\ \lambda a_2 & b_2 & c_2 \\ \lambda a_3 & b_3 & c_3 \end{vmatrix} = \lambda \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

To prove this property, it is sufficient to note that by formula (2) the determinant can be expressed as a sum whose every term contains one element of each row and of each column as a factor.

5°. *If all elements of a column or of a row of a determinant are zero, then the determinant itself is zero.*

This property follows from the preceding property (for $\lambda = 0$).

6°. *If the elements of two columns or of two rows of a determinant are proportional, then the determinant is zero.*

Indeed, if the elements of two columns of a determinant are proportional, then, according to Property 4°, the column factor of the elements of these columns can be put before the sign of the determinant, as a result, a determinant is obtained which has two similar columns and is equal to zero according to Property 3°. For example,

$$\begin{vmatrix} 8 & 4 & 7 \\ 6 & 3 & 9 \\ 4 & 2 & 11 \end{vmatrix} = 2 \begin{vmatrix} 4 & 4 & 7 \\ 3 & 3 & 9 \\ 2 & 2 & 11 \end{vmatrix}.$$

7°. *If every element of the nth column (nth row) of a determinant is the sum of two terms, then the determinant can be represented as the sum of two determinants, one of which has, in the nth column (nth row), the first of the indicated terms and the other has the second terms, the elements which occupy all the other places being the same in all three determinants.* For example,

$$\begin{vmatrix} a_1'+a_1'' & b_1 & c_1 \\ a_2'+a_2'' & b_2 & c_2 \\ a_3'+a_3'' & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1' & b_1 & c_1 \\ a_2' & b_2 & c_2 \\ a_3' & b_3 & c_3 \end{vmatrix} + \begin{vmatrix} a_1'' & b_1 & c_1 \\ a_2'' & b_2 & c_2 \\ a_3'' & b_3 & c_3 \end{vmatrix}.$$

To prove this property, it is sufficient to apply formula (2) to the determinants appearing on the left-hand and right-hand sides of the relation and make sure that the expressions obtained are equal.

8°. *If the elements of a column (a row) of a determinant are summed up with the elements of another column (row) multiplied by any common factor $\lambda$, then the value of the determinant remains unchanged.*

Indeed, according to Property 7°, the determinant resulting from such an addition can be divided into the sum of two determinants, the first of which coincides with the original determinant and the second has two proportional columns and, by virtue of Property 6°, is equal to zero. For example,

$$\begin{vmatrix} a_1+\lambda b_1 & b_1 & c_1 \\ a_2+\lambda b_2 & b_2 & c_2 \\ a_3+\lambda b_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} + \begin{vmatrix} \lambda b_1 & b_1 & c_1 \\ \lambda b_2 & b_2 & c_2 \\ \lambda b_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

To formulate the next property of a determinant, we must get acquainted with the concepts of a minor and of an algebraic adjunct, or cofactor.

The *minor* of an element of a determinant is a determinant obtained from the given determinant by striking out a row and a column in which the element lies. For instance, the minor of the element $a_1$ of the determinant $\Delta$ is the second-order determinant $\begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix}$, the minor of the element $b_1$ is the second-order determinant $\begin{vmatrix} a_2 & c_2 \\ a_3 & c_3 \end{vmatrix}$ and so on.

The *algebraic adjunct*, or *cofactor*, of an element of a determinant is the minor of that element multiplied by $(-1)^p$, where $p$ is the sum of the numbers of the column and the row in which the element lies. The cofactor of an element is denoted by the same capital letter as the element itself. Thus the cofactor of the element $a_1$ is designated as $A_1$, that of the element $b_1$ as $B_1$ and so on.

If, for instance, the element $a_2$ is in the first column and the second row, then $p = 1 + 2 = 3$ for it and the cofactor is

$$A_2 = (-1)^3 \cdot \begin{vmatrix} b_1 & c_1 \\ b_3 & c_3 \end{vmatrix} = b_3c_1 - b_1c_3.$$

Thus the cofactor and the minor of the same element differ only in sign.

9°. *A determinant is equal to the sum of the products of the elements of a column or of a row by their cofactors.*

In other words, the following equalities hold true:

$$\Delta = a_1A_1 + a_2A_2 + a_3A_3, \quad \Delta = a_1A_1 + b_1B_1 + c_1C_1; \tag{3}$$

$$\Delta = b_1B_1 + b_2B_2 + b_3B_3, \quad \Delta = a_2A_2 + b_2B_2 + c_2C_2; \tag{4}$$

$$\Delta = c_1C_1 + c_2C_2 + c_3C_3, \quad \Delta = a_3A_3 + b_3B_3 + c_3C_3. \tag{5}$$

To prove the first equality, for instance, it is sufficient to write the right-hand side of formula (2) in the form

$$\Delta = a_1(b_2c_3 - b_3c_2) + a_2(b_3c_1 - b_1c_3) + a_3(b_1c_2 - b_2c_1).$$

The quantities in parentheses are the cofactors of the elements $a_1$, $a_2$, $a_3$, i.e.

$$b_2c_3 - b_3c_2 = A_1, \quad b_3c_1 - b_1c_3 = A_2, \quad b_1c_2 - b_2c_1 = A_3.$$

From this and from the preceding relation we obtain

$$\Delta = a_1A_1 + a_2A_2 + a_3A_3,$$

and this is what we wished to prove. Formulas (3)-(5) can be proved by analogy.

Writing a determinant with the use of one of the formulas (3)-(5) is its *expansion* by the elements of a column or of a row (the first formula yields the expansion by the elements of the first column etc).

**Example.** Calculate the determinant

$$\Delta = \begin{vmatrix} 2 & 4 & 6 \\ 5 & 12 & 19 \\ 3 & 9 & 17 \end{vmatrix}$$

expanding it by the elements of the first row.

*Solution.* We have

$$\Delta = 2\begin{vmatrix} 12 & 19 \\ 9 & 17 \end{vmatrix} - 4\begin{vmatrix} 5 & 19 \\ 3 & 17 \end{vmatrix} + 6\begin{vmatrix} 5 & 12 \\ 3 & 9 \end{vmatrix} = 8.$$

10°. *The sum of the products of the elements in a column or in a row of a determinant by the cofactors of the respective elements in another column or row is zero.*

We shall prove, for instance, that the sum of the products of the elements of the second column by the respective cofactors of the elements of the first column is zero. For that purpose, we expand determinant (1) by the elements of the first column

$$\Delta = a_1A_1 + a_2A_2 + a_3A_3. \tag{6}$$

The cofactors $A_1$, $A_2$, $A_3$ do not depend on the elements $a_1$, $a_2$, $a_3$ themselves and therefore, if we replace the numbers $a_1$, $a_2$, $a_3$ in both sides of relation (6) by the arbitrary numbers $h_1$, $h_2$, $h_3$, then we get a true equality

$$\begin{vmatrix} h_1 & b_1 & c_1 \\ h_2 & b_2 & c_2 \\ h_3 & b_3 & c_3 \end{vmatrix} = h_1A_1 + h_2A_2 + h_3A_3. \tag{7}$$

Let us now take the elements $b_1$, $b_2$, $b_3$ of the second column as $h_1$, $h_2$, $h_3$ in relation (7). If we take into account that according to Property 3° the determinant with two identical columns is zero, then we obtain

$$b_1A_1 + b_2A_2 + b_3A_3 = 0,$$

and this is what we wished to prove.

By analogy we can prove the equalities

$$c_1A_1 + c_2A_2 + c_3A_3 = 0,$$

$$a_1B_1 + a_2B_2 + a_3B_3 = 0, \quad a_1C_1 + a_2C_2 + a_3C_3 = 0,$$

$$c_1B_1 + c_2B_2 + c_3B_3 = 0, \quad b_1C_1 + b_2C_2 + b_3C_3 = 0$$

and six similar equalities relating to rows rather than to columns:

$$a_1A_3 + b_1B_3 + c_1C_3 = 0, \quad a_2A_3 + b_2B_3 + c_2C_3 = 0,$$

$$a_1A_2 + b_1B_2 + c_1C_2 = 0, \quad a_3A_2 + b_3B_2 + c_3C_2 = 0,$$

$$a_2A_1 + b_2B_1 + c_2C_1 = 0, \quad a_3A_1 + b_3B_1 + c_3C_1 = 0.$$

### 2.10.3. Investigating the Behaviour of a System of Three First-Degree Equations with Three Unknowns

The theory of matrices and determinants has a wide use both in mathematics itself and in its applications. It is a mathematical apparatus which is very convenient and is frequently used in various investigations.

We shall consider the use of matrices and determinants in the investigation of the behaviour of a system of three first-degree equa-

tions with three unknowns $x$, $y$, $z$:

$$\begin{cases} a_1x + b_1y + c_1z = h_1, \\ a_2x + b_2y + c_2z = h_2, \\ a_3x + b_3y + c_3z = h_3 \end{cases} \tag{1}$$

(we assume the coefficients $a_1$, $a_2$, $a_3$, $b_1$, $b_2$, $b_3$, $c_1$, $c_2$, $c_3$ and the constant terms $h_1$, $h_2$, $h_3$ to be known).

The triple of numbers $x_0$, $y_0$, $z_0$ is a *solution of system* (1) if, as a result of the substitution of these numbers for $x$, $y$, $z$, all three equations (1) turn into identities.

In what follows, the following four determinants play the principal part:

$$\Delta = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}, \quad \Delta_x = \begin{vmatrix} h_1 & b_1 & c_1 \\ h_2 & b_2 & c_2 \\ h_3 & b_3 & c_3 \end{vmatrix},$$

$$\Delta_y = \begin{vmatrix} a_1 & h_1 & c_1 \\ a_2 & h_2 & c_2 \\ a_3 & h_3 & c_3 \end{vmatrix}, \quad \Delta_z = \begin{vmatrix} a_1 & b_1 & h_1 \\ a_2 & b_2 & h_2 \\ a_3 & b_3 & h_3 \end{vmatrix}.$$

The determinant $\Delta$ is the *determinant of system* (1). The determinants $\Delta_x$, $\Delta_y$, $\Delta_z$ result from the determinant $\Delta$ of the system when the elements of the first, the second and the third column are replaced by the respective constant terms.

We shall consider two cases separately: (1) when the determinant $\Delta$ of the system is nonzero, and (2) when the determinant is zero.

**Case 1.** $\Delta \neq 0$. We shall prove that a solution of system (1) exists and is unique. For that purpose, we multiply both sides of the first equation of system (1) by the cofactor $A_1$, those of the second equation by $A_2$ and those of the third equation by $A_3$ and then add the resulting equations together. We obtain

$$(a_1A_1 + a_2A_2 + a_3A_3)\,x + (b_1A_1 + b_2A_2 + b_3A_3)\,y$$
$$+ (c_1A_1 + c_2A_2 + c_3A_3)\,z = h_1A_1 + h_2A_2 + h_3A_3.$$

From this, on the basis of Properties 9° and 10° of determinants, we have

$$\Delta \cdot x = h_1A_1 + h_2A_2 + h_3A_3 = \Delta_x.$$

By analogy we find that

$$\Delta \cdot y = h_1B_1 + h_2B_2 + h_3B_3 = \Delta_y,$$
$$\Delta \cdot z = h_1C_1 + h_2C_2 + h_3C_3 = \Delta_z.$$

Thus, from system (1) we have got a system of equations

$$x=\frac{\Delta_x}{\Delta},\quad y=\frac{\Delta_y}{\Delta},\quad z=\frac{\Delta_z}{\Delta}. \tag{2}$$

Formulas (2) are known as *Cramer's* rule.*

To prove that a solution of system (1) exists, we substitute for $x$, $y$, $z$ the values defined by Cramer's rule and make sure that all three equations (1) turn into identities. Let us verify, for instance, that the first equation becomes an identity. We have

$$\begin{aligned} a_1x+b_1y+c_1z &= a_1\frac{\Delta_x}{\Delta}+b_1\frac{\Delta_y}{\Delta}+c_1\frac{\Delta_z}{\Delta}=\frac{1}{\Delta}[a_1(h_1A_1+h_2A_2+h_3A_3)\\ &+b_1(h_1B_1+h_2B_2+h_3B_3)+c_1(h_1C_1+h_2C_2+h_3C_3)]\\ &=\frac{1}{\Delta}[h_1(a_1A_1+b_1B_1+c_1C_1)+h_2(a_1A_2+b_1B_2+c_1C_2)\\ &+h_3(a_1A_3+b_1B_3+c_1C_3)]=h_1, \end{aligned}$$

since, according to Property 9° of determinants,

$$a_1A_1+b_1B_1+c_1C_1=\Delta,$$

and, according to Property 10°,

$$a_1A_2+b_1B_2+c_1C_2=0,\quad a_1A_3+b_1B_3+c_1C_3=0.$$

We have thus established the fact that the first equation of system (1) turns into identity. We can show by analogy that the second and third equations of the system become identities.

Thus a solution of system (1) exists.

Cramer's rule (2) proves that the solution of system (1) is unique since system (2) is a consequence of system (1), and therefore every solution of system (1) is a solution of system (2), i.e. can be expressed by Cramer's formulas.

All this allows us to make the following conclusion: *if the determinant $\Delta$ of system (1) is nonzero, then there is a solution of the system that can be expressed by Cramer's formulas, and that solution is unique.*

**Example.** Find all solutions of the system

$$\begin{cases} x+2y+z=4,\\ 3x-5y+3z=1,\\ 2x+7y-z=8. \end{cases}$$

*Solution.* Since $\Delta=33\neq 0$, this system has a unique solution which is defined by formulas (2):

$$x=\frac{\Delta_x}{\Delta}=\frac{33}{33}=1,\quad y=\frac{\Delta_y}{\Delta}=\frac{33}{33}=1,\quad z=\frac{\Delta_z}{\Delta}=\frac{33}{33}=1.$$

Consequently, $x=1$, $y=1$, $z=1$ is the solution of the system.

* Cramer, Gabriel (1704-1752), a Swiss mathematician.

**Case 2.** $\Delta = 0$. Let at least one of the determinants $\Delta_x$, $\Delta_y$, $\Delta_z$ be nonzero. Then at least one of the formulas (2) is impossible*, i.e. system (2) has no solution, and therefore system (1) has no solution either since system (2) is a consequence of system (1).

Let now $\Delta_x$, $\Delta_y$, $\Delta_z$ be zero. We shall first investigate homogeneous systems.

A *homogeneous system* of three first-degree equations with three unknowns is a system of the form

$$\begin{cases} a_1x + b_1y + c_1z = 0, \\ a_2x + b_2y + c_2z = 0, \\ a_3x + b_3y + c_3z = 0. \end{cases} \tag{3}$$

This system evidently always has a *zero solution*, i.e. $x = 0$, $y = 0$, $z = 0$. If $\Delta \neq 0$, then this solution is unique (by virtue of case 1).

We shall prove that if the determinant $\Delta = 0$, then system (3) has infinitely many nonzero solutions. We shall carry out the proof in two stages.

(1) Assume that at least one of the minors of the determinant $\Delta$ is nonzero. Let $\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \neq 0$, for example. Then we can represent the homogeneous system composed of the first two equations (3) as

$$\begin{cases} a_1x + b_1y = -c_1z, \\ a_2x + b_2y = -c_2z. \end{cases} \tag{4}$$

Let us consider the cofactors $A_3$, $B_3$, $C_3$ of the elements of the third row of the determinant:

$$A_3 = \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix}, \quad B_3 = -\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}, \quad C_3 = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}.$$

Since $C_3 \neq 0$ by the hypothesis, there is a unique solution of system (4) for every $z$. We can write it as $x = \frac{A_3}{C_3} z$ and $y = \frac{B_3}{C_3} z$. We set $z = C_3t$, where $t$ can assume any values. Then, evidently, the homogeneous system (4) has infinitely many solutions defined by formulas

$$x = A_3t, \quad y = B_3t, \quad z = C_3t, \tag{5}$$

where $t$ is an arbitrary number.

It remains to show that $x$, $y$, $z$ defined by formulas (5) turn the third equation of the homogeneous system (3) into identity. Indeed, substituting the expressions for $x$, $y$, $z$ from formulas (5) into the left-hand side of the third equation, we find that

$$a_3x + b_3y + c_3z = (a_3A_3 + b_3B_3 + c_3C_3)\, t = \Delta \cdot t = 0 \cdot t = 0.$$

* Indeed, let $\Delta_x \neq 0$, for example. Then the equality $\Delta \cdot x = \Delta_x$ is impossible since its left-hand side $\Delta \cdot x = 0$ for any $x$ and its right-hand side $\Delta_x \neq 0$.

Thus formulas (5) define the solution of the homogeneous system (3) for any $t$.

(2) Let us assume now that all minors of the determinant $\Delta$ are zero. This means that the coefficients of all three equations of system (3) are proportional. But then the second and the third equation are consequences of the first and can be neglected, and one equation with three unknowns $a_1x + b_1y + c_1z = 0$ evidently has infinitely many solutions (we can assign arbitrary values to two unknowns and find the third one from the equation).

We have thus proved that the homogeneous system (3) with determinant $\Delta$ equal to zero has infinitely many solutions.

We shall consider now system (1) when all four determinants $\Delta$, $\Delta_x$, $\Delta_y$, $\Delta_z$ are zero. We shall prove that if, in this case, system (1) has at least one solution, then it has infinitely many solutions.

Let system (1) have a solution $x_0$, $y_0$, $z_0$. Then the following identities hold true:

$$\begin{cases} a_1x_0 + b_1y_0 + c_1z_0 = h_1, \\ a_2x_0 + b_2y_0 + c_2z_0 = h_2, \\ a_3x_0 + b_3y_0 + c_3z_0 = h_3. \end{cases} \tag{6}$$

Subtracting identities (6) from equations (1) term by term, we get a system of equations equivalent to (1):

$$\begin{cases} a_1(x - x_0) + b_1(y - y_0) + c_1(z - z_0) = 0, \\ a_2(x - x_0) + b_2(y - y_0) + c_2(z - z_0) = 0, \\ a_3(x - x_0) + b_3(y - y_0) + c_3(z - z_0) = 0. \end{cases} \tag{7}$$

This is a homogeneous system of three first-degree equations with unknowns $(x - x_0)$, $(y - y_0)$, $(z - z_0)$ and a determinant $\Delta$ equal to zero. But, in accordance with what we have just proved, this system has infinitely many solutions and, consequently, system (1) also has infinitely many solutions. For instance, when the minor $C_3$ is nonzero, we can represent the solution of system (1), by virtue of formulas (5), in the form

$$x = x_0 + A_3t, \quad y = y_0 + B_3t, \quad z = z_0 + C_3t,$$

where $t$ assumes any values. We have thus proved the assertion and can infer that if $\Delta = \Delta_x = \Delta_y = \Delta_z = 0$, *then system* (1) *either has no solution or has infinitely many solutions.*

By way of example, the reader is invited to consider the following three systems:

$$\begin{cases} x + y + z = 2, \\ 3x + 2y + 2z = 1, \\ 4x + 3y + 3z = 4, \end{cases} \quad \begin{cases} x + y + z = 1, \\ 2x + y + z = 2, \\ 3x + 2y + 2z = 3, \end{cases} \quad \begin{cases} x + y + z = 1, \\ 2x + 2y + 2z = 3, \\ 3x + 3y + 3z = 4 \end{cases}$$

and ascertain that the first system has no solution ($\Delta = 0$, $\Delta_y = 1 \neq 0$), the second system has infinitely many solutions ($\Delta = \Delta_x = \Delta_y = \Delta_z = 0$) which are defined by the formulas $x = 1$, $y = t$, $z = -t$, and the third system has no solution ($\Delta = \Delta_x = \Delta_y = \Delta_z = 0$), but already the first two equations of the system are inconsistent since if we multiply the first equation by 2 and subtract it from the second, we get an impossible equality $0 = 1$.

### 2.10.4. The Matrix Notation of a System of Linear Equations. The Inverse of a Matrix

Let us consider again the system of equations (1) from 2.10.3

$$\begin{cases} a_1x + b_1y + c_1z = h_1, \\ a_2x + b_2y + c_2z = h_2, \\ a_3x + b_3y + c_3z = h_3. \end{cases} \tag{1}$$

We introduce the following designations:

$$A = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}, \quad X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad H = \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix}. \tag{2}$$

Then, using the rule of multiplication of matrices, we can write system (1) in the equivalent matrix form

$$AX = H, \tag{3}$$

where $A$ is the given matrix, $H$ is the given column vector and $X$ is the unknown column vector. The solution of equation (3) is the column vector $X$ which turns equation (3) into an identity.

Let the determinant $\Delta$ of the matrix $A$ be nonzero. Then, as was established in 2.10.3, system (1), and, consequently, system (3), have a unique solution which can be found by means of Cramer's rule. We shall give now another notation of the solution of equation (3), for which purpose we introduce the concept of the inverse of a matrix.

**Definition.** *The inverse of the matrix $A$ is a matrix (designated as $A^{-1}$), which satisfies the conditions*

$$A^{-1} \cdot A = A \cdot A^{-1} = I, \tag{4}$$

*where $I$ is an identity matrix.*

We shall show that if the determinant $\Delta \neq 0$, then the following matrix is the inverse of the matrix $A$:

$$A^{-1} = \begin{pmatrix} A_1/\Delta & A_2/\Delta & A_3/\Delta \\ B_1/\Delta & B_2/\Delta & B_3/\Delta \\ C_1/\Delta & C_2/\Delta & C_3/\Delta \end{pmatrix}. \tag{5}$$

In this matrix, as before, $A_i$, $B_i$, $C_i$ are the cofactors of the respective elements $a_i$, $b_i$, $c_i$ ($i = 1, 2, 3$). We must first prove that matrix (5) satisfies conditions (4). We shall verify, for instance, the validity of the equality

$$AA^{-1} = I.$$

Multiplying the matrix $A$ by the matrix $A^{-1}$ in accordance with the rule of matrix multiplication, we obtain

$$AA^{-1} = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} \begin{pmatrix} A_1/\Delta & A_2/\Delta & A_3/\Delta \\ B_1/\Delta & B_2/\Delta & B_3/\Delta \\ C_1/\Delta & C_2/\Delta & C_3/\Delta \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = I.$$

We have used here the fact that, by virtue of Properties 9° and 10° of a determinant, the following equalities hold true:

$$a_iA_j + b_iB_j + c_iC_j = \begin{cases} \Delta & \text{for } i = j, \\ 0 & \text{for } i \neq j. \end{cases}$$

Therefore, in particular, the element of the matrix $AA^{-1}$, which is in the first row and the first column, is equal to $(a_1A_1 + b_1B_1 + c_1C_1)/\Delta = \Delta/\Delta = 1$, and the element, which is in the first row and the second column, is equal to $(a_1A_2 + b_1B_2 + c_1C_2)/\Delta = 0$.

We can make sure that the other elements of the matrix $AA^{-1}$ are equal to the respective elements of the matrix $I$. Thus $AA^{-1} = I$. We can prove by analogy that $A^{-1}A = I$.

Thus the matrix $A^{-1}$ defined by formula (5) is the inverse of the matrix $A$. It follows from relations (4) that the matrix $A$ is the inverse of the matrix $A^{-1}$. Therefore the matrices $A$ and $A^{-1}$ are said to be *mutually inverse*.

**Remark.** If the determinant of the matrix $A$ is zero ($\Delta = 0$), then there is no inverse matrix.

We shall make use of the inverse matrix to solve equation (3). Premultiplying equation (3) by the matrix $A^{-1}$, we get

$$A^{-1}AX = A^{-1}H. \tag{6}$$

Since $A^{-1}A = I$ and $IX = X$, formula (6) yields a formula

$$X = A^{-1}H. \tag{7}$$

It is easy to verify that the expression obtained for $X$ is indeed a solution of equation (3). In fact, substituting this expression into

equation (3), we have

$$AA^{-1}H = H.$$

This equality is an identity since $AA^{-1} = I$ and $IH = H$.

Thus, if $\Delta \neq 0$, then the solution of equation (3), and, hence, of system (1), can be written in the matrix form (7). Of course, this solution is the same as that obtained in 2.10.3 by means of Cramer's rule. This fact, which follows from the uniqueness of the solution of system (1) for $\Delta \neq 0$ can be immediately verified by substituting expression (5) for $A^{-1}$ and expressions (2) for $X$ and $H$ into formula (7). Then

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} A_1/\Delta & A_2/\Delta & A_3/\Delta \\ B_1/\Delta & B_2/\Delta & B_3/\Delta \\ C_1/\Delta & C_2/\Delta & C_3/\Delta \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix},$$

whence $x = (h_1A_1 + h_2A_2 + h_3A_3)/\Delta = \Delta_x/\Delta$, $y = (h_1B_1 + h_2B_2 + h_3B_3)/\Delta = \Delta_y/\Delta$, $z = (h_1C_1 + h_2C_2 + h_3C_3)/\Delta = \Delta_z/\Delta$, i.e. we have obtained Cramer's formulas.

**Example.** Solve a system of equations

$$\begin{cases} x + 2y + z = 1, \\ 2x + y + z = -1, \\ x + 3y + z = 2. \end{cases}$$

*Solution.* We have

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \\ 1 & 3 & 1 \end{pmatrix}, \quad H = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}.$$

The determinant of the matrix $A$ is $1 \neq 0$ and, consequently, the matrix $A$ has its inverse. From formula (5) we find that

$$A^{-1} = \begin{pmatrix} -2 & 1 & 1 \\ -1 & 0 & 1 \\ 5 & -1 & -3 \end{pmatrix}.$$

Using the matrix $A^{-1}$, we find from formula (7) that

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -2 & 1 & 1 \\ -1 & 0 & 1 \\ 5 & -1 & -3 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix},$$

whence $x = -1$, $y = 1$, $z = 0$.

It should be pointed out in conclusion that, when investigating systems of first-degree equations with many unknowns and in many other problems of mathematics and its applications, we have to deal

with matrices and determinants of an arbitrary order $n$ ($n = 2, 3, 4, 5, \ldots$). The theory of matrices and determinants of an arbitrary order is analogous in structure to the theory of third-order matrices and determinants we have discussed. But its strict construction requires the introduction of additional concepts and the proof of a number of complicated theorems. Those who want to extend their knowledge can get acquainted with this theory and with the theory of first-degree equations with many unknowns using any course in higher algebra.*

**Chapter 11**

# THE CONCEPT, THE LIMIT AND THE CONTINUITY OF A FUNCTION OF SEVERAL VARIABLES

## 2.11.1. The Concept of a Function of Several Variables

**1. Introductory remarks.** Up till now we considered functions of one variable, i.e. functions whose values depend on the values of one independent variable.

When discussing various problems of natural sciences, we have to deal with such relationships between variable quantities in which the numerical values of one of them are completely defined by the values of some other quantities. Thus, for instance, the body temperature at a certain moment $t$ may vary from point to point. Every point of the body is defined by three coordinates $x$, $y$, $z$, and therefore the temperature depends on three variables $x$, $y$, $z$, and if we also take into account the dependence of the temperature on time $t$, then its values will be defined by the values of four variables $x$, $y$, $z$, $t$. The area of a rectangle with sides $x$ and $y$ long is defined by the values of two variables $x$ and $y$, and the volume of a right parallelepiped with edges $x$, $y$, $z$ long is defined by the values of three variables $x$, $y$, $z$. We can cite many examples of this kind of dependence.

This part of the course is devoted to this kind of relationship. For our purpose, we introduce the concept of a function of several variables and develop an apparatus for investigating the behaviour of such functions.

**2. Definition of a function of two and more variables.** We introduce the concept of a function of two variables by analogy with that of a function of one variable.

---

* See, for instance, Kurosh, A.G., *Higher Algebra*, Mir Publishers, Moscow, 1984.

**Definition.** *Let $X$, $Y$, $Z$ be number sets. A function of two variables is the set $f$ of ordered triples of numbers $(x, y, z)$* such that $x \in X$, $y \in Y$, $z \in Z$ and each ordered pair of numbers $(x, y)$ appears in one and only one triple of the set and each $z$ appears in at least one triple. In this case we say that there is a correspondence between the ordered pair of numbers $(x, y)$ and the number $z$ and write $z = f(x, y)$. The number $z$ is the value of the function $f$ at the point $(x, y)$. The variable $z$ is a dependent variable and the variables $x$ and $y$ are independent variables, or arguments, the set $\{(x, y)\}$ is the domain of definition of the function and the set $Z$ is the set of values, or the range, of the function.*

A function of two variables can also be designated as $z = z(x, y)$, $z = \varphi(x, y)$, $z = h(x, y)$, $z = F(x, y)$ etc.

Since, for a fixed rectangular system of coordinates, every ordered pair of numbers $(x, y)$ is associated with a single point $M$ of the plane and, conversely, every point $M$ is associated with a single ordered pair of numbers $(x, y)$, a function of two variables can be regarded as a function of the point $M$ and we can write $z = f(M)$ instead of $z = f(x, y)$. In this case the domain of definition of the function is a set $\{M\}$ of points of the plane. In what follows we shall use these two designations of a function of two variables.

As in the case of one variable, the methods of definition of a function of two variables can differ. In examples we usually use the analytical method of definition and define a function by a formula. In that case the domain of definition of a function is the set of all points of the plane for which the formula has sense.

Here are some examples of functions of two variables.

1. $z = x^2 + y^2$. The domain of definition of this function is the set $\{M\}$ of all pairs of numbers $(x, y)$, i.e. the whole $xy$-plane, and the range is the interval $Z = [0, +\infty)$.

2. $z = \sqrt{1 - x^2 - y^2}$. The domain of definition of this function is the set of all points for which the expression $\sqrt{1 - x^2 - y^2}$ is defined, i.e. the set of all points for which $1 - x^2 - y^2 \geqslant 0$, or $x^2 + y^2 \leqslant 1$. The set of all such points forms a circle with centre at the origin and radius equal to unity, and the range of the function is the interval $[0, 1]$.

3. $z = 1/\sqrt{x^2 + y^2 - 1}$. The domain of definition of this function is the set of points whose coordinates satisfy the inequality $x^2 + y^2 > 1$, i.e. the set of points which lie outside the circle with radius 1 and centre at the origin, and the range is the interval $(0, +\infty)$.

It follows from the examples we have considered that the domain of definition of a function of two variables may be the entire $xy$-plane or its part.

* Recall that the triple of numbers $x$, $y$, $z$ is said to be ordered if it is indicated which of the numbers is the first, which is the second and which is the third.

It is known from analytic geometry that the set of all ordered triples of numbers $(x, y, z)$ forms a coordinate space, and every triple $(x, y, z)$ of numbers of space is associated with a point $M(x, y, z)$ and vice versa. If we take the set $\{M\}$ of points of space instead of the set $\{M\}$ of points of the plane, we can similarly give a definition of a function of three variables $u = f(M)$ or $u = f(x, y, z)$. The domain of definition of a function of three variables is the whole space or its part. Thus, for instance, the function $u = x^2 + y^2 + z^2$ is defined in the whole space and the function $u = \ln xyz$ is defined on the set of points of space whose coordinates satisfy the inequality $xyz > 0$. In the former case the range of the function is the interval $[0, +\infty)$ and in the latter case it is the interval $(-\infty, +\infty)$.

By analogy we can give a definition of a function of four variables $u = f(x, y, z, t)$. In this case the set of ordered quadruples of numbers $(x, y, z, t)$ forms a so-called four-dimensional space and each quadruple $(x, y, z, t)$ is a point of that space. The domain of definition of a function of four variables, however, no longer has a visual geometric interpretation. By analogy we can give a definition of a function of five variables and, in general, of $n$ variables $u = f(x_1, x_2, \ldots, x_n)$.

In what follows, we consider in detail functions of two variables. It should be remembered that the generalization of the definitions and results obtained to functions of three and more variables does not bring about any essential differences.

### 2.11.2. Geometric Representation of a Function of Two Variables

It is known that a function of one variable is represented in the plane as a curve defined by the equation $y = f(x)$. A function of two variables is represented in space as a surface which is defined by the equation $z = f(x, y)$, i.e. the formula which defines the function is itself an equation of the surface.

Analytic geometry deals with various surfaces and their equations. Thus, for instance, the equation $z - 2x + 5y + 10 = 0$ is the equation of a plane. This plane is the graph of the function $z = 2x - 5y - 10$.

The equation $x^2 + y^2 + z^2 = R^2$ is the equation of a sphere of radius $R$ with centre at the origin. On the other hand, a sphere is the union of the graphs of two functions $z = \sqrt{R^2 - x^2 - y^2}$ and $z = -\sqrt{R^2 - x^2 - y^2}$.

In many cases it is difficult to construct the graphs of functions of two variables and therefore there is one more method of representing a function of two variables which is based on the section of the surface $z = f(x, y)$ by the planes $z = c$, where $c$ is any number, i.e. by the planes which are parallel to the $xy$-plane.

The *level* (or *contour*) *line* of the function $z = f(x, y)$ is the set of points $(x, y)$ of the $xy$-plane at which the function assumes the same value $c$. For different $c$ we evidently obtain different level lines for a given function.

If we take numbers $c_1, c_2, \ldots, c_n$ which form an arithmetic progression with a common difference $h$, then we get a number of level lines whose mutual positions give a notion of the graph of the function, i.e. of the shape of the surface. In the places where the level lines are

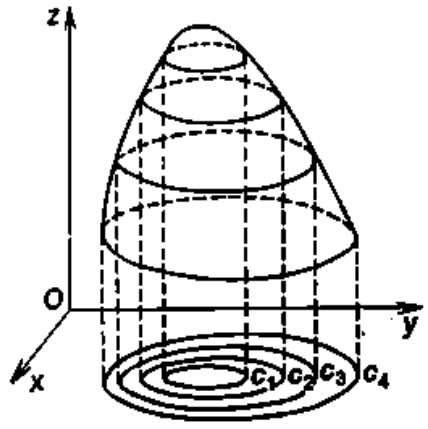



Fig. 162

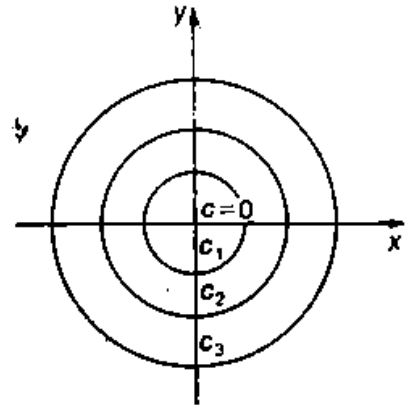



Fig. 163

"denser", the function varies more quickly (the surface is steeper), and in the places where the level lines are rarer, the function varies more slowly (the slope of the surface is less) (Fig. 162). Clearly, the smaller $h$, the more complete is the notion of the graph of a function.

The term "level lines" is taken from cartography where level lines are lines at which the altitude of the points of the earth surface above the sea level is constant. Using these lines, we can form an opinion not only of the altitude of a point above the sea level but also of the relief of the locality, which is especially important when the locality is mountainous.

**Example.** Construct the level lines of the function $z = x^2 + y^2$.

*Solution.* The level lines of this function are defined by the equation $x^2 + y^2 = c$ $(0 \leqslant c < +\infty)$. Assigning different values to $c$, we get a family of level lines which are concentric circles. For $c = 0$ the circle degenerates into a point (0, 0) (Fig. 163).

Since in this case the level lines are circles with centres at the origin, the graph of the function must be a surface of revolution about the $z$-axis. Indeed, it is known from analytic geometry that the equation $z = x^2 + y^2$ defines a paraboloid of revolution.

### 2.11.3. The Limit of a Function of Two Variables

We introduce the concept of the δ-neighbourhood of the point $M_0(x_0, y_0)$ and the concept of the convergent sequence of points of the plane.

**Definition 1.** *The set $\{M(x, y)\}$ of all points whose coordinates $x$ and $y$ satisfy the inequality $\sqrt{(x-x_0)^2+(y-y_0)^2} < \delta$, or, briefly, $\rho(M, M_0) < \delta$ is the δ-neighbourhood of the point $M_0(x_0, y_0)$.*

To put it otherwise, the δ-neighbourhood of the point $M_0$ is composed of all points which lie inside a circle with centre $M_0$ of radius δ.

Let us consider a sequence of points $M_1(x_1, y_1)$, $M_2(x_2, y_2)$, ..., $M_n(x_n, y_n)$, .... We shall use the brief notation $\{M_n\}$*.

**Definition 2.** *The sequence $\{M_n\}$ of points is said to converge to the point $M_0$ if for any $\varepsilon > 0$ there is a number $N$ such that the inequality $\rho(M_n, M_0) < \varepsilon$ is satisfied for $n > N$. In this case the point $M_0$ is the limit of the sequence $\{M_n\}$.*

The notation is $\lim\limits_{n\to\infty} M_n = M_0$, or $M_n \to M_0$ as $n \to \infty$.

Note that the concept of a convergent sequence of points of the plane is a generalization of the concept of a convergent number sequence. Indeed, the specifying of the sequence $\{M_n\}$ of points on a straight line is equivalent to the specifying of the number sequence $\{x_n\}$ and in this case inequality $\rho(M_n, M_0) < \varepsilon$ turns into an inequality $|x_n - x_0| < \varepsilon$.

We shall now define the limit of a function of two variables. Its definition is similar to that of the limit of a function of one variable.**

Let the function $z = f(M)$ be defined on a set $\{M\}$ and assume that the point $M_0 \in \{M\}$ or $M_0 \notin \{M\}$ but possesses the property that in any δ-neighbourhood of this point there is at least one point of the set $\{M\}$ different from $M_0$.

**Definition 3.** *A number $A$ is the limit of the function $z = f(M)$ at a point $M_0$ if for any sequence of points $M_1, M_2, \ldots, M_n, \ldots$ ($M_n \neq M_0$, $M_n \in \{M\}$) converging to $M_0$ the corresponding sequence of the values of the function $f(M_1), f(M_2), \ldots, f(M_n), \ldots$ converges to $A$.*

The notation is $\lim\limits_{M\to M_0} f(M) = A$, or $\lim\limits_{\substack{x\to x_0\\ y\to y_0}} f(x, y) = A$.

Thus, for instance, the function $f(x, y) = x^2 + y^2$ is defined on the whole plane. Let us find the limit of this function at the point $M_0(1, 2)$. For any sequence $\{M_n\}$ of points, which converges to the

* The definition of the sequence $\{M_n\}$ is equivalent to the definition of two number sequences $\{x_n\}$ and $\{y_n\}$ since a point on the plane is defined by two coordinates $x$ and $y$.

** We recommend the reader to read 1.4.2, Item 1, again.

point $M_0$, we have

$$\lim_{n\to\infty}(x_n^2+y_n^2)=\lim_{n\to\infty}x_n^2+\lim_{n\to\infty}y_n^2=1^2+2^2=5$$

and consequently, $\lim\limits_{\substack{x\to 1\\ y\to 2}}(x^2+y^2)=5$.

Here is an example of a function which has no limit at a point. The function $f(x, y) = (x - y)/(x + y)$ is defined everywhere except for the points of the straight line $x + y = 0$. We shall show that it has no limit at the point (0, 0). For that purpose, we choose two sequences of points $M_n$ $(1/n, 0)$ and $M'_n$ $(0, 1/n)$ converging to the point (0, 0) and then

$$\lim_{n\to\infty}f(M_n)=\lim_{n\to\infty}\frac{1/n-0}{1/n+0}=1,\quad \lim_{n\to\infty}f(M'_n)=\lim_{n\to\infty}\frac{0-1/n}{0+1/n}=-1.$$

Thus two different sequences of points converging to the origin are associated with two sequences of the values of the function, which have different limits. Consequently, according to Definition 3, this function has no limit at the point (0, 0).

We have given the definition of the limit of a function of two variables using the concept of the limit of a sequence. As for a function of one variable, we can give an equivalent definition using the "ε-δ" terminology.

**Definition 4.** *A number $A$ is the limit of the function $z = f(M)$ at a point $M_0$ if for any $\varepsilon > 0$ there is $\delta > 0$ such that the inequality $|f(M) - A| < \varepsilon$ holds true for all points $M \in \{M\}$ which satisfy the condition $0 < \rho(M, M_0) < \delta$.*

Using the logical symbols, we can write this definition as

$$(\forall\varepsilon>0)\ (\exists\delta>0)\ (\forall M\in\{M\},\ 0<\rho(M,M_0)<\delta):$$
$$|f(M)-A|<\varepsilon.$$

We can carry out the proof of the equivalence of Definitions 3 and 4 in the same way as we did in the case of a function of one variable, except that in the proof of Theorem 4.1 we must replace the sequence $\{x_n\}$ by the sequence $\{M_n\}$ of points, the point $x_0$ by the point $M_0$, the differences $|x - x_0|$ and $|x_n - x_0|$ by the distances $\rho(M, M_0)$ and $\rho(M_n, M_0)$ and the number sequence $\{f(x_n)\}$ by the number sequence $\{f(M_n)\}$ respectively.

Using the definition of the limit of a function of two variables, we can extend the fundamental theorems on the limits of functions of one variable to functions of two variables. For instance, consider the following theorem.

**Theorem.** *Assume that the functions $f(M)$ and $g(M)$ are defined on the same set $\{M\}$ and have limits $B$ and $C$ at a point $M_0$. Then the*

*functions* $f(M) \pm g(M)$, $f(M) \cdot g(M)$ *and* $\frac{f(M)}{g(M)}$ $(C \neq 0)$ *have limits at the point* $M_0$ *which are equal to* $B \pm C$, $B \cdot C$ *and* $\frac{B}{C}$ *respectively.*

The proof of this theorem is similar to the proof of Theorem 4.3 and can be obtained from it by the formal replacement of the letters $x$ and $x_0$ by the letters $M$ and $M_0$, only in that case we must use Definition 3 of the limit of a function of two variables rather than Definition 1 of the limit of a function of one variable.

**Definition 5.** *The function* $z = f(M)$ *is infinitesimal at a point* $M = M_0$ *(or as* $M \to M_0$*) if* $\lim\limits_{M \to M_0} f(M) = 0$.

If the function $z = f(M)$ has a limit equal to $A$ at the point $M_0$, then the function $\alpha(M) = f(M) - A$ is infinitesimal at the point $M_0$. Indeed, $\lim\limits_{M \to M_0} \alpha(M) = \lim\limits_{M \to M_0} [f(M) - A] = \lim\limits_{M \to M_0} f(M) - \lim\limits_{M \to M_0} A = A - A = 0$. From this we get a special representation of the function which has a limit equal to $A$ at the point $M_0$: $f(M) = A + \alpha(M)$, where $\lim\limits_{M \to M_0} \alpha(M) = 0$. In this case we say that in the neighbourhood of the point $M_0$ the function $f(M)$ differs from the number $A$ by an infinitesimal function.

We can compare infinitesimal functions of two variables in the same way as we compare infinitesimal functions of one variable, and, as in the case of one variable, the symbol $o(\beta)$ signifies any infinitesimal (at the given point $M_0$) function of a higher order of smallness than the infinitesimal (at the point $M_0$) function $\beta(M)$, i.e. $\lim\limits_{M \to M_0} \frac{o(\beta)}{\beta} = 0$.

### 2.11.4. The Continuity of a Function of Two Variables

We introduce the concept of the continuity of a function of two variables using the concept of a limit.

**1. Definition of the continuity of a function of two variables.** Assume that the function $f(M)$ is defined on a set $\{M\}$, the point $M_0 \in \{M\}$, and any $\delta$-neighbourhood of the point $M_0$ contains points of the set $\{M\}$.

**Definition 1.** *The function* $z = f(M)$ *is continuous at a point* $M_0$ *if the limit of the function at this point exists and is equal to the value of the function at this point, i.e.*

$$\lim_{M \to M_0} f(M) = f(M_0), \quad or \quad \lim_{\substack{x \to x_0 \\ y \to y_0}} f(x, y) = f(x_0, y_0).$$

In accordance with the definition of the limit of a function in the terms of sequences, this definition of the continuity of a function at

the point $M_0$ is equivalent to the fact that, for any sequence $\{M_n\}$ $(M_n \in \{M\})$ such that $\lim\limits_{n\to\infty} M_n = M_0$, the sequence $\{f(M_n)\}$ converges and

$$\lim_{n\to\infty} f(M_n) = f(M_0), \quad \text{or} \quad \lim_{n\to\infty} f(x_n, y_n) = f(x_0, y_0).$$

The points at which a function does not possess the property of continuity are called the *points of discontinuity* of the function. For example, the function

$$f(x, y) = \begin{cases} \dfrac{x-y}{x+y} & \text{for } x+y \neq 0, \\ 1 & \text{for } x+y=0 \end{cases}$$

is discontinuous at the point (0, 0) since the limit of this function does not exist as $x \to 0$ and $y \to 0$ and the function

$$f(x, y) = \begin{cases} x^2+y^2 & \text{everywhere except for } x=1 \text{ and } y=2, \\ 0 & \text{for } x=1 \text{ and } y=2 \end{cases}$$

is discontinuous at the point (1, 2) since $\lim\limits_{\substack{x\to 1\\ y\to 2}} f(x, y) = 5$ and $f(1, 2) = 0$.

We shall formulate the definition of the continuity of a function using the definition of the limit of a function in the "ε-δ" terms.

**Definition 2.** *The function $z = f(M)$ is continuous at a point $M_0$ if for any $\varepsilon > 0$ there is $\delta > 0$ such that the inequality $|f(M) - f(M_0)| < \varepsilon$ holds true for all points $M \in \{M\}$ which satisfy the condition $\rho(M, M_0) < \delta$.*

Using the logical symbols, we can write Definition 2 as

$$(\forall \varepsilon > 0)\,(\exists \delta > 0)\,(\forall M \in \{M\},\ \rho(M, M_0) < \delta):\ |f(M) - f(M_0)| < \varepsilon.$$

As for a function of one variable, we can use these definitions of the continuity and the corresponding theorems on limits and prove that the arithmetic operations involving continuous functions and the construction of composite functions from continuous functions lead to continuous functions.

In what follows, we use Definition 1 of the continuity of a function written in a different form.

The *total increment* of the function $z = f(M)$ at the point $M_0$ is the function $\Delta z$ defined by the formula

$$\Delta z = f(M) - f(M_0),$$

where $M$ is any point of the domain of definition of the function. Let the points $M_0$ and $M$ have the coordinates $(x_0, y_0)$ and $(x, y)$ respectively. We designate $x - x_0 = \Delta x$ and $y - y_0 = \Delta y$. Using

these designations, we get the following expression for $\Delta z$:

$$\Delta z = f(x_0 + \Delta x,\ y_0 + \Delta y) - f(x_0,\ y_0).$$

**Definition 3.** *The function $z = f(M)$ is continuous at a point $M_0$ if its total increment at this point is an infinitesimal function as $M \to M_0$, i.e.*

$$\lim_{M \to M_0} \Delta z = \lim_{M \to M_0} [f(M) - f(M_0)] = 0, \text{ or } \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \Delta z = 0.$$

This condition is evidently equivalent to the condition $\lim\limits_{M \to M_0} f(M) = f(M_0)$ from Definition 1.

**Example.** The function $z = x^2 + y^2$ is continuous at any point $(x, y)$. Indeed, the total increment of this function at the point $(x, y)$ has the form

$$\Delta z = [(x + \Delta x)^2 + (y + \Delta y)^2] - (x^2 + y^2)$$
$$= 2x\,\Delta x + 2y\,\Delta y + (\Delta x)^2 + (\Delta y)^2.$$

It is evident that $\Delta z \to 0$ as $\Delta x \to 0$ and $\Delta y \to 0$, i.e., according to Definition 3, the function $z = x^2 + y^2$ is continuous at the point $(x, y)$.

The function $z = f(M)$ is continuous on a set $\{M\}$ if it is continuous at every point of this set.

**2. Principal properties of continuous functions of two variables.** We give, without proof, the principal properties of continuous functions of two variables since they are similar, in the main, to the corresponding properties of functions of one variable. We shall first introduce some concepts for the sets $\{M\}$ of points of the plane.

**Definition 4.** *The set $\{M\}$ of points of the plane is said to be connected if any two points of the set can be connected by a solid line consisting of points of this set.*

For instance, a circle is a connected set, whereas a set consisting of two circles which have no points in common is not connected.

**Definition 5.** *A point $M$ is an interior point of the set $\{M\}$ if there is a $\delta$-neighbourhood of the point which consists of points of this set.*

**Definition 6.** *The set $\{M\}$ consisting only of interior points is an open set.*

**Definition 7.** *A connected open set $\{M\}$ of points is an open domain, or, briefly, a domain.*

The interior of a triangle, of a circle, of an ellipse etc. is a simple domain.

**Definition 8.** *A point $M$ is a boundary point of a domain if in any of its $\delta$-neighbourhoods there are points which belong and do not belong*

*to that domain. The set of all boundary points of a domain is the boundary of the domain.*

For instance, for a domain consisting of points which lie inside a circle the boundary is the circumference.

**Definition 9.** *The set $\{M\}$ of points formed by a domain and its boundary is a closed domain.*

**Definition 10.** *The set $\{M\}$ is bounded if there is a circle inside which it lies.*

A line segment and a triangle are bounded sets. A straight line is not a bounded set.

A bounded closed domain in which a function of two variables is defined is an analogue of a line segment for a function of one variable.

We shall now formulate the principal properties of continuous functions of two variables.

1°. If the function $z = f(M)$ is continuous in a bounded closed domain, it is bounded in that domain, i.e. there is a number $k$ such that the inequality $|f(M)| \leqslant k$ is satisfied for all points of the domain.

2°. If the function $z = f(M)$ is continuous in a bounded closed domain, then it attains its least upper bound and greatest lower bound in that domain.

3°. If the function $z = f(M)$ is continuous in a domain, then it assumes all intermediate values between any two of its values, i.e. if $A < C < B$, where $A$ and $B$ are some values of the function $f(M)$ in that domain, then there is a point $M_0$ in the domain at which $f(M_0) = C$.

From this it follows, in particular, that if $M_1$ and $M_2$ are points of the given domain and $f(M_1) < 0$ and $f(M_2) > 0$, then there is a point $M_0$ in the domain at which $f(M_0) = 0$.

4°. If the function $z = f(M)$ is continuous in a bounded closed domain, then it is uniformly continuous in that domain, i.e. for any $\varepsilon > 0$ there is $\delta > 0$ such that the inequality $|f(M'') - f(M')| < \varepsilon$ holds true for any two points $M'$ and $M''$ of the domain which satisfy the condition $\rho(M', M'') < \delta$.

It should be pointed out in conclusion that the concepts of the limit, the continuity and the enumerated properties of functions of two variables can be easily generalized to functions of three and more variables.

Chapter 12

# PARTIAL DERIVATIVES AND THE DIFFERENTIABILITY OF FUNCTIONS OF SEVERAL VARIABLES

## 2.12.1. Partial Derivatives

Let the function $z = f(M)$ be defined in a neighbourhood of the point $M(x, y)$*. We assign an arbitrary increment $\Delta x$ to the variable $x$ at the point $M$, leaving the value of the variable $y$ unchanged, i.e. pass from the point $M(x, y)$ to the point $M_1(x + \Delta x, y)$ on the plane. In this case $\Delta x$ is such that the point $M_1$ lies in the indicated neighbourhood of the point $M$. Then the corresponding increment of the function

$$\Delta_x z = f(x + \Delta x, y) - f(x, y)$$

is a *partial increment of the function with respect to the variable* $x$ at the point $M(x, y)$.

By analogy we can define the *partial increment of the function with respect to the variable* $y$:

$$\Delta_y z = f(x, y + \Delta y) - f(x, y).$$

**Definition.** *If there is a limit*

$$\lim_{\Delta x \to 0} \frac{\Delta_x z}{\Delta x} \left( \lim_{\Delta y \to 0} \frac{\Delta_y z}{\Delta y} \right),$$

*then it is a partial derivative of the function* $z = f(M)$ *at the point* $M$ *with respect to the variable* $x$ *(with respect to the variable* $y$*) and may be denoted by one of the following symbols*:

$$z'_x,\ f'_x,\ \frac{\partial f}{\partial x},\ \frac{\partial z}{\partial x} \left( z'_y,\ f'_y,\ \frac{\partial f}{\partial y},\ \frac{\partial z}{\partial y} \right)$$

(sometimes the symbols for partial derivatives are written without a prime).

It follows from the definition that the partial derivative of a function of two variables with respect to the variable $x$ is an ordinary derivative of a function of one variable $x$ for a fixed value of the variable $y$. Therefore partial derivatives can be calculated by means of formulas and laws of calculation of the derivatives of functions of one variable.

* A neighbourhood of the point $M$ is an arbitrary domain which contains the point $M$ in its interior. In particular, if the domain is the interior of a circle of radius $\delta$ with centre at the point $M$, then the neighbourhood is a $\delta$-neighbourhood of the point $M$.

**Examples.**

1. $z = x^2 - 2xy^2 + y^3$, $\frac{\partial z}{\partial x} = 2x - 2y^2$, $\frac{\partial z}{\partial y} = -4xy + 3y^2$.

2. $z = \arctan \frac{x}{y}$, $\frac{\partial z}{\partial x} = \frac{y}{x^2+y^2}$, $\frac{\partial z}{\partial y} = \frac{-x}{x^2+y^2}$.

3. $z = x^2 \sin y$, $\frac{\partial z}{\partial x} = 2x \sin y$, $\frac{\partial z}{\partial y} = x^2 \cos y$.

4. $z = xye^{x+2y}$, $\frac{\partial z}{\partial x} = e^{x+2y}y(1+x)$, $\frac{\partial z}{\partial y} = e^{x+2y}x(1+2y)$.

Note that we have defined partial derivatives of the function $z = f(x, y)$ at the point $M$ in whose neighbourhood the function is defined, i.e. at an interior point of the domain of definition of the function. If $M(x, y)$ is a boundary point of the domain of definition of the function, then $\Delta_x z$ ($\Delta_y z$) may not be defined since the point $M_1(x + \Delta x, y)$ ($M_2(x, y + \Delta y)$) may not belong to the domain of definition of the function for any $\Delta x \neq 0$ ($\Delta y \neq 0$). This occurs, for instance, for the point $M_0$ shown in Fig. 164. In this case, if there is a partial derivative $z'_x$ at the interior points $M$ of the domain and there is a limit $\lim\limits_{M \to M_0} z'_x(M)$, then we assume, by definition, that

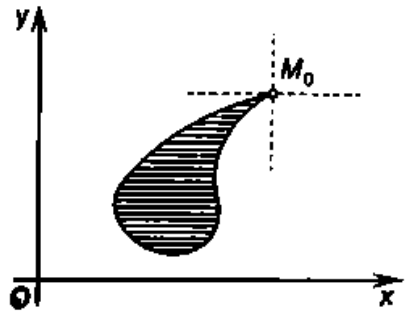


Fig. 164

$$z'_x(M_0) = \lim_{M \to M_0} z'_x(M).$$

We can similarly define $z'_y(M_0)$.

### 2.12.2. The Differentiability of a Function

**1. Definition of differentiability.** Recall that the total increment of the function $z = f(M)$ at the point $M(x, y)$, which corresponds to the increments $\Delta x$ and $\Delta y$ of the variables $x$ and $y$, is the function

$$\Delta z = f(x + \Delta x, y + \Delta y) - f(x, y).$$

Let the function $z = f(M)$ be defined in a neighbourhood of the point $M$.

**Definition.** *The function $z = f(M)$ is differentiable at the point $M$ if its total increment at this point can be represented as*

$$\Delta z = A\,\Delta x + B\,\Delta y + \alpha(\Delta x, \Delta y)\,\Delta x + \beta(\Delta x, \Delta y)\,\Delta y, \quad (1)$$

*where $A$ and $B$ are some numbers independent of $\Delta x$ and $\Delta y$, and $\alpha(\Delta x, \Delta y)$ and $\beta(\Delta x, \Delta y)$ are infinitesimal functions as $\Delta x \to 0$ and $\Delta y \to 0$.*

It is known that if a function of one variable is differentiable at a point, then it is continuous and has a derivative at that point. From the existence of a derivative of a function of one variable at a given point it follows that the function is differentiable at that point. Let us see how these properties can be extended to functions of two variables.

**2. The necessary conditions for differentiability.**

**Theorem 12.1.** *If the function $z = f(M)$ is differentiable at the point $M$, then it is continuous at this point.*

*Proof.* If the function $z = f(M)$ is differentiable at the point $M$, then, as follows from relation (1), $\lim\limits_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \Delta z = 0$, and this means that the function is continuous at the point $M$. ■

**Theorem 12.2.** *If the function $z = f(M)$ is differentiable at the point $M(x, y)$, then it has partial derivatives $f'_x(x, y)$ and $f'_y(x, y)$ at that point and*

$$f'_x(x, y) = A, \quad f'_y(x, y) = B.$$

*Proof.* Since the function $z = f(M)$ is differentiable at the point $M$, relation (1) holds true. Setting $\Delta y = 0$, we have $\Delta_x z = A\,\Delta x + \alpha(\Delta x, 0)\,\Delta x$, where $\alpha(\Delta x, 0)$ is an infinitesimal function as $\Delta x \to 0$. Dividing $\Delta_x z = A\,\Delta x + \alpha(\Delta x, 0)\,\Delta x$ by $\Delta x$ and passing to the limit as $\Delta x \to 0$, we obtain

$$\lim_{\Delta x \to 0} \frac{\Delta_x z}{\Delta x} = \lim_{\Delta x \to 0} [A + \alpha(\Delta x, 0)] = A.$$

Consequently, there is a partial derivative $f'_x(x, y) = A$ at the point $M$.

We can show by analogy that there is also a partial derivative $f'_y(x, y) = B$ at the point $M$. ■

Statements converse to Theorems 12.1 and 12.2 are not true, i.e. the continuity of a function of two variables at the point $M$ and the existence of its partial derivatives at this point do not imply its differentiability.

For instance, the function $f(x, y) = \sqrt{x^2 + y^2}$ is continuous at the point (0, 0) but does not have partial derivatives at this point. Indeed,

$$\frac{f(0+\Delta x, 0) - f(0, 0)}{\Delta x} = \frac{\sqrt{(0+\Delta x)^2 + 0} - 0}{\Delta x} = \frac{|\Delta x|}{\Delta x}.$$

But the function $\frac{|\Delta x|}{\Delta x}$ does not have a limit as $\Delta x \to 0$ and, consequently, $f'_x(0, 0)$ does not exist.

We can prove by analogy that $f'_y(0, 0)$ does not exist either. Since this function does not possess partial derivatives at the point (0, 0), it is not differentiable at this point.

The function

$$f(x, y) = \begin{cases} 0 \text{ on the coordinate axes,} \\ 1 \text{ at the other points of the plane} \end{cases}$$

has partial derivatives with respect to $x$ and $y$ at the point (0, 0). This follows from the fact that $f(x, 0) \equiv 0$ and $f(0, y) \equiv 0$ and, consequently, $f'_x(0, 0) = 0$ and $f'_y(0, 0) = 0$. But $f(x, y)$ is not continuous at this point since along a straight line, for instance, $y = x$ $\lim\limits_{\substack{x\to 0\\ y\to 0}} f(x, y) = 1$ and $f(0, 0) = 0$. Consequently, $f(x, y)$ is not differentiable at the point (0, 0).

And the last example. The function $f(x, y) = \sqrt{|x|\,|y|}$ is continuous at the point (0, 0) since $f(0, 0) = 0$ and $\lim\limits_{\substack{x\to 0\\ y\to 0}} f(x, y) = 0$ and has partial derivatives with respect to $x$ and $y$ at this point:

$$f'_x(0, 0) = \lim_{\Delta x\to 0} \frac{\sqrt{|\Delta x|\cdot 0} - 0}{\Delta x} = 0,$$

$$f'_y(0, 0) = \lim_{\Delta y\to 0} \frac{\sqrt{0\cdot|\Delta y|} - 0}{\Delta y} = 0,$$

but nevertheless this function is not differentiable at the point (0, 0). Indeed, the total increment of the function at the point (0, 0) is

$$\Delta z = \sqrt{|\Delta x|\,|\Delta y|}.$$

If the function were differentiable at the point (0, 0), then, as follows from relation (1) and Theorem 12.2, there would hold an equality

$$\Delta z = 0\cdot\Delta x + 0\cdot\Delta y + \alpha\,\Delta x + \beta\,\Delta y,$$

where $\alpha$ and $\beta$ tend to zero as $\Delta x \to 0$ and $\Delta y \to 0$. Furthermore, from the relation

$$\alpha\,\Delta x + \beta\,\Delta y = \left(\frac{\alpha\,\Delta x + \beta\,\Delta y}{\rho}\right)\rho = \gamma\rho, \text{ where } \rho = \sqrt{(\Delta x)^2 + (\Delta y)^2},$$

we find that $\gamma \to 0$ as $\rho \to 0$ (see the derivation of formula (3) in 2.12.4). But in this case, for any $\Delta x = \Delta y$,

$$\gamma = \frac{\sqrt{|\Delta x|\,|\Delta y|}}{\sqrt{(\Delta x)^2 + (\Delta y)^2}} = \frac{1}{\sqrt{2}},$$

i.e. it is not an infinitesimal function as $\rho \to 0$.

Thus the function $f(x, y) = \sqrt{|x|\,|y|}$ is continuous at the point (0, 0), has partial derivatives at this point, and nevertheless is not differentiable at this point.

**3. The sufficient conditions for differentiability.**

**Theorem 12.3.** *If the function $z = f(M)$ has partial derivatives at some $\delta$-neighbourhood of the point $M$ and those derivatives are continuous at the point $M$ itself, then the function is differentiable at the point $M$.*

*Proof.* We assign to the variables $x$ and $y$ some increments $\Delta x$ and $\Delta y$ so small that the point $M_1(x + \Delta x, y + \Delta y)$ does not leave the bounds of this $\delta$-neighbourhood of the point $M$. The total increment of the function

$$\Delta z = f(x + \Delta x, y + \Delta y) - f(x, y)$$

can be written as

$$\Delta z = [f(x + \Delta x, y + \Delta y) - f(x, y + \Delta y)] + [f(x, y + \Delta y) - f(x, y)]. \qquad (2)$$

We can regard the expression $[f(x + \Delta x, y + \Delta y) - f(x, y + \Delta y)]$ as an increment of the function $f(x, y + \Delta y)$ of one variable $x$ (the second argument has a constant value equal to $y + \Delta y$). Since, by the hypothesis, this function has a derivative $f'_x(x, y + \Delta y)$, we find, in accordance with Lagrange's theorem, that

$$f(x + \Delta x, y + \Delta y) - f(x, y + \Delta y) = f'_x(x + \theta_1 \Delta x, y + \Delta y)\,\Delta x, \quad 0 < \theta_1 < 1.$$

Reasoning by analogy, we have, for the expression $[f(x, y + \Delta y) - f(x, y)]$, a relation

$$f(x, y + \Delta y) - f(x, y) = f'_y(x, y + \theta_2 \Delta y)\,\Delta y, \quad 0 < \theta_2 < 1.$$

The derivatives $f'_x$ and $f'_y$ are continuous at the point $M(x, y)$ and therefore

$$\lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} f'_x(x + \theta_1 \Delta x, y + \Delta y) = f'_x(x, y),$$

$$\lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} f'_y(x, y + \theta_2 \Delta y) = f'_y(x, y).$$

From this it follows that

$$f'_x(x + \theta_1 \Delta x, y + \Delta y) = f'_x(x, y) + \alpha(\Delta x, \Delta y),$$
$$f'_y(x, y + \theta_2 \Delta y) = f'_y(x, y) + \beta(\Delta x, \Delta y),$$

where $\alpha(\Delta x, \Delta y)$ and $\beta(\Delta x, \Delta y)$ are infinitesimal functions as $\Delta x \to 0$ and $\Delta y \to 0$. Substituting the expressions obtained into formula (2) for $\Delta z$, we find that

$$\Delta z = f'_x(x, y)\,\Delta x + f'_y(x, y)\,\Delta y + \alpha(\Delta x, \Delta y)\,\Delta x + \beta(\Delta x, \Delta y)\,\Delta y,$$

and this precisely means that the function $z = f(M)$ is differentiable at the point $M$. ■

**Corollary.** *The continuity of partial derivatives implies the continuity of the function itself.*

Theorem 12.3 is very significant for establishing the differentiability of a function since a direct verification of the differentiability of a function by means of the definition is often difficult, whereas verification of continuity of partial derivatives proves to be much easier.

We want to point out in conclusion that the concept of differentiability for functions of three and more variables can be introduced by analogy with the case of a function of two variables.

### 2.12.3. Derivatives of Composite Functions

Let $z = f(x, y)$ be a function of two variables $x$ and $y$, each of which, in its turn, is a function of the independent variable $t$: $x = x(t)$ and $y = y(t)$. Then the function $z = f[x(t), y(t)]$ is a composite function of the independent variable $t$, and the variables $x$ and $y$ are intermediate variables.

Consider the following theorem.

**Theorem 12.4.** *If the functions $x = x(t)$ and $y = y(t)$ are differentiable at the point $t$, and the function $z = f(x, y)$ is differentiable at the point $M[x(t), y(t)]$, then the composite function $z = f[x(t), y(t)]$ is also differentiable at the point $t$. The derivative of this composite function can be found from the formula*

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}. \tag{1}$$

*Proof.* We assign an arbitrary increment $\Delta t$ to the variable $t$, and then the functions $x(t)$ and $y(t)$ get increments $\Delta x$ and $\Delta y$ respectively and the function $z = f(x, y)$, in its turn, gets an increment

$$\Delta z = f(x + \Delta x, y + \Delta y) - f(x, y).$$

Since the function $z = f(x, y)$ is differentiable at the point $M(x, y)$ where $x = x(t)$ and $y = y(t)$, we can write $\Delta z$ as

$$\Delta z = f'_x(x, y)\,\Delta x + f'_y(x, y)\,\Delta y + \alpha(\Delta x, \Delta y)\,\Delta x + \beta(\Delta x, \Delta y)\,\Delta y,$$

where $\alpha(\Delta x, \Delta y)$ and $\beta(\Delta x, \Delta y)$ are infinitesimal functions as $\Delta x \to 0$ and $\Delta y \to 0$. We complete the definition of these functions for $\Delta x = 0$ and $\Delta y = 0$, setting $\alpha(0, 0) = 0$ and $\beta(0, 0) = 0$.

Dividing both sides of the equality for $\Delta z$ by $\Delta t$, we obtain

$$\frac{\Delta z}{\Delta t} = f'_x(x, y)\frac{\Delta x}{\Delta t} + f'_y(x, y)\frac{\Delta y}{\Delta t} + \alpha(\Delta x, \Delta y)\frac{\Delta x}{\Delta t} + \beta(\Delta x, \Delta y)\frac{\Delta y}{\Delta t}. \tag{2}$$

By the hypothesis, $\lim\limits_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt}$ and $\lim\limits_{\Delta t \to 0} \frac{\Delta y}{\Delta t} = \frac{dy}{dt}$. In addition, since the functions $x(t)$ and $y(t)$ are differentiable at the point $t$, they are continuous at this point, i.e. $\Delta x \to 0$ and $\Delta y \to 0$ as $\Delta t \to 0$ and, as a consequence, $\alpha(\Delta x, \Delta y) \to 0$ and $\beta(\Delta x, \Delta y) \to 0$. Therefore the terms $\alpha(\Delta x, \Delta y)\frac{\Delta x}{\Delta t}$ and $\beta(\Delta x, \Delta y)\frac{\Delta y}{\Delta t}$ tend to zero as $\Delta t \to 0$.

We have thus proved that as $\Delta t \to 0$ there is a limit of the right-hand side of relation (2) and, consequently, there is a limit of the left-hand side

$$\lim_{\Delta t \to 0} \frac{\Delta z}{\Delta t} = \frac{dz}{dt}$$

and

$$\frac{dz}{dt} = f'_x(x, y)\frac{dx}{dt} + f'_y(x, y)\frac{dy}{dt}$$

or

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}. \blacksquare$$

**Remark.** Do not mix up the cases when a derivative is designated as "$\partial$" and when as "d".

**Examples.**

1. Let $z = f(x, y)$, $x = t^3 + 2$, $y = 3t^4 - 1$. From formula (1) we have

$$\frac{dz}{dt} = \frac{\partial z}{\partial x} 3t^2 + \frac{\partial z}{\partial y} 12t^3.$$

2. Let $z = x \sin\frac{x}{y}$, $x = 1 + 3t$, $y = \sqrt{1+t^2}$. From formula (1) we have

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$$

$$= \left(\sin\frac{x}{y} + \frac{x}{y}\cos\frac{x}{y}\right) 3 - \frac{x^2}{y^2}\cos\frac{x}{y}\frac{t}{\sqrt{1+t^2}}.$$

3. Let $z = x^2y^3$, $x = t$, $y = t^2$. From formula (1) we get

$$\frac{dz}{dt} = 2xy^3 \cdot 1 + 3x^2y^2 \cdot 2t.$$

Taking into account that $x = t$ and $y = t^2$, we find that $\frac{dz}{dt} = 2t(t^2)^3 + 3t^2(t^2)^2 2t = 8t^7$. On the other hand, we can find $\frac{dz}{dt}$ first expressing $z$ in terms of $t$. We have $z = x^2y^3 = t^2(t^2)^3 = t^8$, whence $\frac{dz}{dt} = 8t^7$, and this obviously coincides with the result obtained from formula (1).

If $z = f(x, y)$, where $y = \varphi(x)$, then $z = f[x, \varphi(x)]$ is a composite function of $x$. On the basis of formula (1), in which $x$ now plays the part of $t$, we obtain

$$\frac{dz}{dx} = \frac{\partial z}{\partial x}\frac{dx}{dx} + \frac{\partial z}{\partial y}\frac{dy}{dx},$$

and since $\frac{dx}{dx} = 1$, we have

$$\frac{dz}{dx} = \frac{\partial z}{\partial x} + \frac{\partial z}{\partial y}\frac{dy}{dx}.$$

We can similarly find the derivative of a composite function when the number of intermediate variables is larger than two. For instance, if $u = f(x, y, z)$, where $x = x(t)$, $y = y(t)$, $z = z(t)$, then formula (1) assumes the form

$$\frac{du}{dt} = \frac{\partial u}{\partial x}\frac{dx}{dt} + \frac{\partial u}{\partial y}\frac{dy}{dt} + \frac{\partial u}{\partial z}\frac{dz}{dt}.$$

We shall consider now a more general case. Let $z = f(x, y)$ be a function of two variables $x$ and $y$, which, in turn, depend on two or more independent variables. For instance, let $x = x(u, v)$ and $y = y(u, v)$. Then the function $z = f[x(u, v), y(u, v)]$ is a composite function of the independent variables $u$ and $v$, and the variables $x$ and $y$ are intermediate.

If the functions $x(u, v)$ and $y(u, v)$ are differentiable at the point $M'(u, v)$ and the function $z = f(x, y)$ is differentiable at the point $M(x, y)$, where $x = x(u, v)$ and $y = y(u, v)$, then the composite function $z = f[x(u, v), y(u, v)]$ is differentiable at the point $M'(u, v)$ and its partial derivatives at this point can be found from the formulas

$$\begin{cases} \dfrac{\partial z}{\partial u} = \dfrac{\partial z}{\partial x}\dfrac{\partial x}{\partial u} + \dfrac{\partial z}{\partial y}\dfrac{\partial y}{\partial u}, \\ \dfrac{\partial z}{\partial v} = \dfrac{\partial z}{\partial x}\dfrac{\partial x}{\partial v} + \dfrac{\partial z}{\partial y}\dfrac{\partial y}{\partial v}. \end{cases} \tag{3}$$

We shall not prove the differentiability of a composite function. To calculate its partial derivatives, we fix the value of one of the variables, $u$ or $v$. Then we find ourselves in the conditions of Theorem 12.4 we have just proved. In both cases formula (1) yields formulas (3).

**Examples.**

1. Let $z = f(x, y)$, $x = u^2 + 2v$, $y = u^2/v$. From formulas (3) we have

$$\frac{\partial z}{\partial u} = \frac{\partial z}{\partial x}2u + \frac{\partial z}{\partial y}\frac{2u}{v}, \quad \frac{\partial z}{\partial v} = \frac{\partial z}{\partial x}2 + \frac{\partial z}{\partial y}\left(-\frac{u^2}{v^2}\right).$$

2. Let $z=x^2y^2$, $x=u+v$, $y=u/v$. From formulas (3) we have

$$\frac{\partial z}{\partial u}=2xy^2\cdot 1+2x^2y\cdot\frac{1}{v},\quad \frac{\partial z}{\partial v}=2xy^2\cdot 1+2x^2y\cdot\left(-\frac{u}{v^2}\right).$$

Substitute the expressions $x=u+v$ and $y=u/v$ into these formulas. Find $\frac{\partial z}{\partial u}$ and $\frac{\partial z}{\partial v}$, first expressing $z$ in terms of $u$ and $v$. Compare the results.

3. Let $z=x^2-y^2$, $x=u\cos v$, $y=u\sin v$. From formulas (3) we have

$$\frac{\partial z}{\partial u}=2x\cos v-2y\sin v,\quad \frac{\partial z}{\partial v}=-2xu\sin v-2yu\cos v.$$

If $z=f(x)$, where $x=x(u, v)$, then $z=f[x(u, v)]$ is a composite function which, via the variable $x$, depends on two variables $u$ and $v$, and its partial derivatives can also be found from formulas (3):

$$\begin{cases}\dfrac{\partial z}{\partial u}=\dfrac{dz}{dx}\dfrac{\partial x}{\partial u},\\[2mm] \dfrac{\partial z}{\partial v}=\dfrac{dz}{dx}\dfrac{\partial x}{\partial v}.\end{cases}$$

Pay attention to the designations of the derivatives in these formulas.

We can generalize formulas (3) to the case of a larger number of intermediate variables.

For instance, if $w=f(x, y, z)$ is a function of three variables $x, y, z$, and each of them depends on $u$ and $v$, then formulas (3) assume the form

$$\begin{cases}\dfrac{\partial w}{\partial u}=\dfrac{\partial w}{\partial x}\dfrac{\partial x}{\partial u}+\dfrac{\partial w}{\partial y}\dfrac{\partial y}{\partial u}+\dfrac{\partial w}{\partial z}\dfrac{\partial z}{\partial u},\\[2mm] \dfrac{\partial w}{\partial v}=\dfrac{\partial w}{\partial x}\dfrac{\partial x}{\partial v}+\dfrac{\partial w}{\partial y}\dfrac{\partial y}{\partial v}+\dfrac{\partial w}{\partial z}\dfrac{\partial z}{\partial v}.\end{cases}$$

### 2.12.4. The Differential of a Function

**1. Definition of a differential.** Recall that if the function $z=f(M)$ is differentiable at the point $M$, then its total increment at this point can be represented as

$$\Delta z=A\,\Delta x+B\,\Delta y+\alpha(\Delta x, \Delta y)\,\Delta x+\beta(\Delta x, \Delta y)\,\Delta y, \qquad (1)$$

where $\alpha(\Delta x, \Delta y)$ and $\beta(\Delta x, \Delta y)$ are infinitesimal functions as $\Delta x\to 0$ and $\Delta y\to 0$.

**Definition.** *The differential* dz *of the function $z=f(M)$ differentiable at the point $M$ is a part of the total increment of this function at the point $M$, which is linear with respect to the increments $\Delta x$ and $\Delta y$,*

*i.e.*

$$dz = A\ \Delta x + B\ \Delta y. \tag{2}$$

Using Theorem 12.2, we can rewrite expression (2) as follows:

$$dz = f'_x(x, y)\ \Delta x + f'_y(x, y)\ \Delta y.$$

The differentials of the independent variables $x$ and $y$ are the increments of these variables, i.e. $dx = \Delta x$ and $dy = \Delta y$. Then we can write the differential of the function as

$$dz = f'_x(x, y)\ dx + f'_y(x, y)\ dy.$$

It follows from relations (1) and (2) that the difference between the total increment and the differential of the function at the point $M$

$$\Delta z - dz = \alpha(\Delta x, \Delta y)\ \Delta x + \beta(\Delta x, \Delta y)\ \Delta y$$

is an infinitesimal, as $\Delta x \to 0$ and $\Delta y \to 0$, of a higher order than that of $\rho = \sqrt{(\Delta x)^2 + (\Delta y)^2}$ ($\rho$ is the distance between the points $M(x, y)$ and $M_1(x + \Delta x, y + \Delta y)$). Indeed,

$$\lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{\Delta z - dz}{\rho} = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \left(\alpha \frac{\Delta x}{\rho} + \beta \frac{\Delta y}{\rho}\right) = 0$$

since $\alpha$ and $\beta$ are infinitesimals, and $\frac{\Delta x}{\rho}$ and $\frac{\Delta y}{\rho}$ are bounded functions: $\left|\frac{\Delta x}{\rho}\right| \leqslant 1$ and $\left|\frac{\Delta y}{\rho}\right| \leqslant 1$. Hence we get $\Delta z - dz = o(\rho)$ or

$$\Delta z = dz + o(\sqrt{(\Delta x)^2 + (\Delta y)^2}). \tag{3}$$

Dropping the quantity $o(\rho)$ for sufficiently small $\Delta x$ and $\Delta y$, we get an approximate formula

$$\Delta z \cong dz,$$

which is widely used in approximations since it is easier to calculate a differential than a total increment.

**2. A tangent plane and a normal to a surface. The geometric meaning of a differential.** Just as the differential of a function of one variable is an increment of the "ordinate of a tangent" in terms of geometry, the differential of a function of two variables is an increment of the "applicate of a tangent plane". We shall introduce the concept of a tangent plane to a surface at the point $N_0$.

*The plane which passes through the point $N_0$ of a surface is known as a tangent plane to the surface at that point if the angle between the secant (straight line), which passes through the point $N_0$ and any point $N$ of the surface, and the plane tends to zero when the point $N$ tends to the point $N_0$.*

Let a surface be defined by the equation $z = f(x, y)$ and the function $f(x, y)$ be differentiable at the point $M_0(x_0, y_0)$.

We shall prove that the tangent plane to the surface at the point $N_0(x_0, y_0, z_0)$, where $z_0 = f(x_0, y_0)$, is defined by the equation

$$z - z_0 = f'_x(x_0, y_0)(x - x_0) + f'_y(x_0, y_0)(y - y_0). \qquad (4)$$

Indeed, it is known from analytic geometry that equation (4) defines the plane which passes through the point $N_0(x_0, y_0, z_0)$ and has a normal vector $\mathbf{n} = \{f'_x, f'_y, -1\}$. To ascertain that this plane is a tangent plane, it is sufficient to prove that the angle $\varphi$ between the vector $\mathbf{n}$ and the vector $\overrightarrow{N_0N}$ of any secant $N_0N$ tends to $\pi/2$ when the point $N$ tends to $N_0$. We designate the coordinates of the point $N$ as $(x, y, z)$, where $z = f(x, y)$. Since the coordinates of the vector $\mathbf{n}$ are $f'_x, f'_y, -1$, and the coordinates of the vector $\overrightarrow{N_0N}$ are $x - x_0$, $y - y_0$, $z - z_0$, we have

$$\cos\varphi = \frac{f'_x(x-x_0)+f'_y(y-y_0)-(z-z_0)}{\sqrt{f'^2_x+f'^2_y+1}\sqrt{(x-x_0)^2+(y-y_0)^2+(z-z_0)^2}}.$$

But, as follows from relation (3), $f'_x(x-x_0)+f'_y(y-y_0)-(z-z_0) = o(\rho)$, where $\rho = \sqrt{(x-x_0)^2+(y-y_0)^2}$, and therefore

$$|\cos\varphi| \leqslant \frac{|o(\rho)|}{\sqrt{(x-x_0)^2+(y-y_0)^2}} = \frac{|o(\rho)|}{\rho} \to 0$$

as $\rho \to 0$. Hence $\lim\limits_{N\to N_0} \varphi = \pi/2$, and this is what we wished to prove.

The normal vector $\mathbf{n} = \{f'_x, f'_y, -1\}$ of the tangent plane is a *normal* to the surface $z = f(x, y)$ at the point $N_0$. We assume that $x - x_0 = \Delta x$, $y - y_0 = \Delta y$, $z - z_0 = \Delta z$, and then we find from relation (4) that the increment $\Delta z$ of the "applicate of the tangent plane" is defined by the formula

$$\Delta z = f'_x(x_0, y_0)\,\Delta x + f'_y(x_0, y_0)\,\Delta y,$$

i.e. it indeed coincides with the differential $dz$ of the function $z = f(x, y)$.

### 2.12.5. The Directional Derivative. The Gradient

Consider a function $z = f(M)$ defined in a neighbourhood of the point $M(x, y)$ and an arbitrary unit vector $\mathbf{l} = \{\cos\alpha, \cos\beta\}$ (Fig. 165).

To characterize the rate of variation of the function at the point $M(x, y)$ in the direction of the vector $\mathbf{l}$, we introduce the concept of a directional derivative. For that purpose, we draw a straight line $L$ through the point $M$ so that one of the directions on it coincides with the direction of the vector $\mathbf{l}$ and take a point $M_1(x + \Delta x,$

$y + \Delta y$) on the directed straight line. We designate the value of the line segment $MM_1$ as $\Delta l$, i.e. $\Delta l = \sqrt{(\Delta x)^2 + (\Delta y)^2}$ if the point $M_1$ is positioned as in Fig. 165, and $\Delta l = -\sqrt{(\Delta x)^2 + (\Delta y)^2}$ if the point $M_1$ lies on the other side of the point $M$. In this case the function $f(M)$ gets an increment

$$\Delta z = f(x + \Delta x,\ y + \Delta y) - f(x,\ y).$$

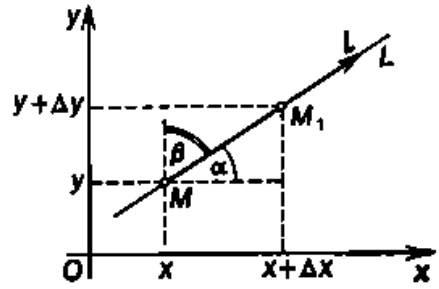


Fig. 165

**Definition 1.** *The limit of the ratio* $\frac{\Delta z}{\Delta l}$ *as* $\Delta l \to 0$ ($M_1 \to M$), *provided that it exists, is the derivative of the function* $z = f(M)$ *at the point* $M(x, y)$ *in the direction of the vector* $\mathbf{l}$ *and is designated as* $\frac{\partial z}{\partial l}$, *i.e.*

$$\lim_{\Delta l \to 0} \frac{\Delta z}{\Delta l} = \frac{\partial z}{\partial l}.$$

We assume now that the function $f(M)$ is differentiable at the point $M$. Then its increment at this point along the straight line $L$ can be written in the form

$$\Delta z = f'_x(x,\ y)\,\Delta x + f'_y(x,\ y)\,\Delta y + \alpha_1(\Delta x,\ \Delta y)\,\Delta x + \beta_1(\Delta x,\ \Delta y)\,\Delta y,$$

where $\alpha_1$ and $\beta_1$ are infinitesimal functions as $\Delta l \to 0$. Dividing both sides of the relation by $\Delta l$ and taking into account that $\Delta x = \Delta l \cos\alpha$ and $\Delta y = \Delta l \sin\alpha = \Delta l \cos\beta$, we obtain

$$\frac{\Delta z}{\Delta l} = f'_x(x,\ y)\cos\alpha + f'_y(x,\ y)\cos\beta + \alpha_1(\Delta x,\ \Delta y)\cos\alpha + \beta_1(\Delta x,\ \Delta y)\cos\beta.$$

Passing to the limit in this relation as $\Delta l \to 0$, we get a formula for a directional derivative

$$\frac{\partial z}{\partial l} = \frac{\partial z}{\partial x}\cos\alpha + \frac{\partial z}{\partial y}\cos\beta. \tag{1}$$

It follows from formula (1) that a directional derivative is a linear combination of partial derivatives, the directional cosines being as if weight factors showing the contribution the corresponding partial derivative makes to the directional derivative.

In particular, $\frac{\partial z}{\partial l} = \frac{\partial z}{\partial x}$ for $\alpha = 0$ and $\beta = \frac{\pi}{2}$, $\frac{\partial z}{\partial l} = \frac{\partial z}{\partial y}$ for $\alpha = \frac{\pi}{2}$ and $\beta = 0$. Hence it follows that the partial derivatives with respect to $x$ and $y$ are special cases of a directional derivative.

**Example.** Calculate the derivative of the function $z = x^2 + y^2x$ at the point $M$ (1, 2) in the direction of the vector $\overrightarrow{MM_1}$, where $M_1$ is a point with coordinates (3, 0).

*Solution.* We find the unit vector $\mathbf{l}$ which has the given direction:

$$\overrightarrow{MM_1} = \{2, \ -2\} = 2\mathbf{i} - 2\mathbf{j}, \quad |\overrightarrow{MM_1}| = 2\sqrt{2},$$

$$\mathbf{l} = \frac{\overrightarrow{MM_1}}{|\overrightarrow{MM_1}|} = \frac{2\mathbf{i} - 2\mathbf{j}}{2\sqrt{2}} = \frac{1}{\sqrt{2}}\mathbf{i} - \frac{1}{\sqrt{2}}\mathbf{j},$$

whence it follows that $\cos\alpha = 1/\sqrt{2}$ and $\cos\beta = -1/\sqrt{2}$. We calculate the partial derivatives of the function at the point $M$ (1, 2): $f'_x(x, y) = 2x + y^2$ and $f'_y(x, y) = 2xy$, whence we get $f'_x(1, 2) = 6$ and $f'_y(1, 2) = 4$. From formula (1) we obtain

$$\frac{\partial z}{\partial l} = 6\frac{1}{\sqrt{2}} - 4\frac{1}{\sqrt{2}} = \sqrt{2}.$$

**Definition 2.** *The gradient of the function $z = f(M)$ at the point $M(x, y)$ is a vector whose coordinates are equal to the corresponding partial derivatives $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$ taken at the point $M(x, y)$.*

The designation is $\operatorname{grad} z = \left\{\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right\}$.

Using the concept of the gradient of a function and taking into account that the vector $\mathbf{l}$ has coordinates $\cos\alpha$ and $\cos\beta$, we represent formula (1) as a scalar product of the vectors grad $z$ and $\mathbf{l}$:

$$\frac{\partial z}{\partial l} = \frac{\partial z}{\partial x}\cos\alpha + \frac{\partial z}{\partial y}\cos\beta = \operatorname{grad} z \cdot \mathbf{l}. \tag{2}$$

On the other hand, from the definition of a scalar product we have

$$\operatorname{grad} z \cdot \mathbf{l} = |\operatorname{grad} z| \cdot |\mathbf{l}| \cos\varphi, \tag{3}$$

where $|\operatorname{grad} z|$ is the length of the vector grad $z$, $\varphi$ is the angle between the vectors $\mathbf{l}$ and grad $z$. Comparing formulas (2) and (3) and taking into account that $|\mathbf{l}| = 1$, we obtain

$$\frac{\partial z}{\partial l} = |\operatorname{grad} z| \cos\varphi.$$

It follows from the last relation that the directional derivative of the function has the greatest value for $\cos\varphi = 1$ ($\varphi = 0$), i.e. when the direction of the vector $\mathbf{l}$ coincides with the direction of grad $z$, and then $\frac{\partial z}{\partial l} = |\operatorname{grad} z|$.

Thus the gradient of the function $z = f(M)$ at the point $M(x, y)$ characterizes the direction and the value of the maximum rate of increase of the function at this point.

We can similarly define the directional derivative for a function of three variables $u = f(x, y, z)$. We derive a formula

$$\frac{\partial u}{\partial l} = \frac{\partial u}{\partial x}\cos\alpha + \frac{\partial u}{\partial y}\cos\beta + \frac{\partial u}{\partial z}\cos\gamma,$$

introduce the concept of a gradient

$$\operatorname{grad} u = \left\{\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial u}{\partial z}\right\}$$

and investigate its properties.

The concepts of a directional derivative and a gradient of a function play an important part in many applications.

### 2.12.6. Higher-Order Partial Derivatives and Differentials

**1. Higher-order partial derivatives.** Assume that the partial derivatives $f'_x(x, y)$ and $f'_y(x, y)$ of the function $z = f(M)$ defined in the neighbourhood of the point $M$ exist at every point of the neighbourhood. In this case the partial derivatives are functions of two variables $x$ and $y$ defined in that neighbourhood of the point $M$. We shall call them *first-order partial derivatives.*

In their turn, the partial derivatives with respect to the variables $x$ and $y$ of the functions $f'_x(x, y)$ and $f'_y(x, y)$ at the point $M$, provided that they exist, are *second-order partial derivatives* of the function $f(M)$ at this point and are designated as follows:

$$\frac{\partial^2 z}{\partial x^2} = f''_{xx}(x, y) = f^{(2)}_{x^2}(x, y), \quad \frac{\partial^2 z}{\partial y\,\partial x} = f''_{yx}(x, y) = f^{(2)}_{yx}(x, y),$$

$$\frac{\partial^2 z}{\partial y^2} = f''_{yy}(x, y) = f^{(2)}_{y^2}(x, y), \quad \frac{\partial^2 z}{\partial x\,\partial y} = f''_{xy}(x, y) = f^{(2)}_{xy}(x, y).$$

Second-order partial derivatives of the form $f''_{yx}(x, y)$ and $f''_{xy}(x, y)$ are *mixed partial derivatives.*

**Examples.**

1. $z = x^4 + 4x^2y^3 + 7xy + 1$. We have

$$\frac{\partial z}{\partial x} = 4x^3 + 8xy^3 + 7y, \quad \frac{\partial z}{\partial y} = 12x^2y^2 + 7x.$$

Consequently,

$$\frac{\partial^2 z}{\partial x^2} = 12x^2 + 8y^3, \quad \frac{\partial^2 z}{\partial x\,\partial y} = \frac{\partial^2 z}{\partial y\,\partial x} = 24xy^2 + 7, \quad \frac{\partial^2 z}{\partial y^2} = 24x^2y.$$

2. $z = \sin x \cos y$. We have

$$\frac{\partial z}{\partial x} = \cos x \cos y, \quad \frac{\partial z}{\partial y} = -\sin x \sin y.$$

Consequently,

$$\frac{\partial^2 z}{\partial x^2} = -\sin x \cos y, \quad \frac{\partial^2 z}{\partial x\, \partial y} = \frac{\partial^2 z}{\partial y\, \partial x} = -\cos x \sin y,$$

$$\frac{\partial^2 z}{\partial y^2} = -\sin x \cos y.$$

In both examples the mixed partial derivatives $f''_{yx}(x, y)$ and $f''_{xy}(x, y)$ are equal. But, in general, the values of mixed derivatives depend on the order of differentiation. Thus, for instance, the function

$$f(x, y) = \begin{cases} xy\,\dfrac{x^2-y^2}{x^2+y^2} & \text{for } x^2+y^2 \neq 0, \\ 0 & \text{for } x^2+y^2=0 \end{cases}$$

has mixed partial derivatives $f''_{yx}(x, y)$ and $f''_{xy}(x, y)$ at the point (0, 0) but they are unequal. Indeed,

$$f'_x(x, y) = \begin{cases} \dfrac{y(x^4-y^4+4x^2y^2)}{(x^2+y^2)^2} & \text{for } x^2+y^2 \neq 0, \\ 0 & \text{for } x^2+y^2=0. \end{cases}$$

Consequently,

$$f''_{yx}(0, 0) = \lim_{\Delta y \to 0} \frac{f'_x(0, 0+\Delta y) - f'_x(0, 0)}{\Delta y} = -1.$$

Carrying out similar calculations, we get $f''_{xy}(0, 0) = 1$. Thus $f''_{yx}(0, 0) \neq f''_{xy}(0, 0)$.

The question may arise concerning the conditions under which the values of mixed derivatives do not depend on the order in which differentiation is carried out. The following theorem answers this question.

**Theorem 12.5.** *If the derivatives $f''_{xy}(x, y)$ and $f''_{yx}(x, y)$ exist in a $\delta$-neighbourhood of the point $M(x, y)$ and are continuous at the point $M$ itself, then they are equal at this point, i.e. there holds an equality*

$$f''_{xy}(x, y) = f''_{yx}(x, y).$$

*Proof.* Consider an expression

$$A = [f(x+\Delta x, y+\Delta y) - f(x+\Delta x, y)] - [f(x, y+\Delta y) - f(x, y)],$$

where $\Delta x$ and $\Delta y$ are any arbitrary numbers so small that the point $M_1(x+\Delta x, y+\Delta y)$ is in that $\delta$-neighbourhood of the point $M$.

We introduce an auxiliary function

$$\varphi(x) = f(x, y+\Delta y) - f(x, y)$$

and then the expression for $A$ can be regarded as an increment of the function $\varphi(x)$ of one variable $x$ differentiable on the interval

$[x, x + \Delta x]$:

$$A = \Delta\varphi = \varphi(x + \Delta x) - \varphi(x).$$

Therefore, applying Lagrange's theorem to this difference, we write

$$A = \Delta\varphi = \varphi'(x + \theta_1 \Delta x)\,\Delta x$$

$$= [f'_x(x + \theta_1 \Delta x, y + \Delta y) - f'_x(x + \theta_1 \Delta x, y)]\,\Delta x, \quad 0 < \theta_1 < 1.$$

The expression in brackets can be regarded as an increment of the function $f'_x(x + \theta_1 \Delta x, y)$ of one variable $y$ differentiable on the interval $[y, y + \Delta y]$. Applying once again Lagrange's theorem (with respect to the vairable $y$), we obtain

$$A = f''_{yx}(x + \theta_1 \Delta x, y + \theta_2 \Delta y)\,\Delta x\,\Delta y, \quad 0 < \theta_1, \quad \theta_2 < 1. \tag{1}$$

On the other hand, if we introduce an auxiliary function

$$\psi(y) = f(x + \Delta x, y) - f(x, y),$$

then, by analogy, we get

$$A = \Delta\psi = \psi(y + \Delta y) - \psi(y),$$

and then

$$A = f''_{xy}(x + \theta_4 \Delta x, y + \theta_3 \Delta y)\,\Delta y\,\Delta x, \quad 0 < \theta_3, \quad \theta_4 < 1. \tag{2}$$

Comparing (1) and (2), we obtain

$$f''_{yx}(x + \theta_1 \Delta x, y + \theta_2 \Delta y) = f''_{xy}(x + \theta_4 \Delta x, y + \theta_3 \Delta y).$$

Passing to the limit in this relation as $\Delta x \to 0$ and $\Delta y \to 0$ and taking the continuity of the partial derivatives $f''_{yx}(x, y)$ and $f''_{xy}(x, y)$ at the point $M$ into account, we get

$$\lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} f''_{yx}(x + \theta_1 \Delta x, y + \theta_2 \Delta y) = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} f''_{xy}(x + \theta_4 \Delta x, y + \theta_3 \Delta y)$$

or

$$f''_{xy}(x, y) = f''_{yx}(x, y). \quad \blacksquare$$

We can introduce the partial derivatives of the third, fourth, . . ., $n$th order by analogy with the second-order partial derivatives and prove a theorem of type 12.5 on the equality of mixed derivatives of any order.

**2. Differentials of higher orders.** In 2.12.4 we introduced the concept of the differential of the function $z = f(M)$ differentiable at the point $M$ and got a formula

$$dz = f'_x(x, y)\,dx + f'_y(x, y)\,dy. \tag{3}$$

We shall call $dz$ a *first-order differential*. For convenience, we agree to denote differentials not only by the symbol d but also by $\delta$ (say, $\delta x$, $\delta y$).

Let the functions $f'_x(x, y)$ and $f'_y(x, y)$ be differentiable at the point $M$. We consider $dx$ and $dy$ in the expression for $dz$ to be cons-

tant factors. Then the function dz is a function of only the variables $x$ and $y$, which is differentiable at the point $M$, and its differential has the form

$$\delta(\mathrm{d}z) = \delta[f'_x(x, y)\,\mathrm{d}x + f'_y(x, y)\,\mathrm{d}y]$$
$$= [f'_x(x, y)\,\mathrm{d}x + f'_y(x, y)\,\mathrm{d}y]'_x\,\delta x + [f'_x(x, y)\,\mathrm{d}x + f'_y(x, y)\,\mathrm{d}y]'_y\,\delta y. \tag{4}$$

The differential $\delta(\mathrm{d}z)$ of the differential dz at the point $M$, taken for $\delta x = \mathrm{d}x$ and $\delta y = \mathrm{d}y$, is the *second-order differential of the function* $z = f(M)$ at the point $M$ and is designated as $\mathrm{d}^2z$. In its turn, the differential $\delta(\mathrm{d}^2z)$ of the differential $\mathrm{d}^2z$, taken for $\delta x = \mathrm{d}x$ and $\delta y = \mathrm{d}y$, is the *third-order differential* of the function $z = f(M)$ and is designated as $\mathrm{d}^3z$, and so on. The differential $\delta(\mathrm{d}^{n-1}z)$ of the differential $\mathrm{d}^{n-1}z$, taken for $\delta x = \mathrm{d}x$ and $\delta y = \mathrm{d}y$, is the *nth-order differential* (or the *nth differential*) of the function $z = f(M)$ and is designated as $\mathrm{d}^nz$.

Thus, for the $n$th differential of the function $z = f(M)$, there holds a formula

$$\mathrm{d}^nz = \delta(\mathrm{d}^{n-1}z)\Big|_{\substack{\delta x=\mathrm{d}x\\ \delta y=\mathrm{d}y}}.$$

When we seek the second (and subsequent) differentials, we usually calculate $\delta(\mathrm{d}z)$ and equate the differentials of the arguments ($\delta x = \mathrm{d}x$ and $\delta y = \mathrm{d}y$) simultaneously.

We shall use formula (4) to find the expression for the second-order differential:

$$\mathrm{d}^2z = \delta(\mathrm{d}z)\Big|_{\substack{\delta x=\mathrm{d}x\\ \delta y=\mathrm{d}y}} = (f'_x\,\mathrm{d}x + f'_y\,\mathrm{d}y)'_x\,\mathrm{d}x + (f'_x\,\mathrm{d}x + f'_y\,\mathrm{d}y)'_y\,\mathrm{d}y$$
$$= f''_{xx}(\mathrm{d}x)^2 + f''_{xy}\,\mathrm{d}x\,\mathrm{d}y + f''_{yx}\,\mathrm{d}y\,\mathrm{d}x + f''_{yy}(\mathrm{d}y)^2.$$

If $f''_{xy}$ and $f''_{yx}$ are continuous, then, according to Theorem 12.5, the terms $f''_{xy}\,\mathrm{d}x\,\mathrm{d}y$ and $f''_{yx}\,\mathrm{d}y\,\mathrm{d}x$ are equal and therefore

$$\mathrm{d}^2z = f''_{xx}(\mathrm{d}x)^2 + 2f''_{xy}\,\mathrm{d}x\,\mathrm{d}y + f''_{yy}(\mathrm{d}y)^2.$$

Similarly

$$\mathrm{d}^3z = f'''_{x^3}(\mathrm{d}x)^3 + 3f'''_{x^2y}(\mathrm{d}x)^2\,\mathrm{d}y + 3f'''_{xy^2}\,\mathrm{d}x(\mathrm{d}y)^2 + f'''_{y^3}(\mathrm{d}y)^3,$$

. . . . . .

$$\mathrm{d}^nz = f^{(n)}_{x^n}(\mathrm{d}x)^n + nf^{(n)}_{x^{n-1}y}(\mathrm{d}x)^{n-1}\,\mathrm{d}y +$$
$$\ldots + \frac{n(n-1)\ \ldots\ (n-k+1)}{k!}\,f^{(n)}_{x^{n-k}y^k}(\mathrm{d}x)^{n-k}(\mathrm{d}y)^k + \ldots + f^{(n)}_{y^n}(\mathrm{d}y)^n.$$

The formula for $\mathrm{d}^nz$ resembles the expansion of the $n$-degree binomial by the binomial formula and therefore the expression for $\mathrm{d}^nz$ can be symbolically written in the form convenient for memorizing:

$$\mathrm{d}^nz = \left(\frac{\partial}{\partial x}\,\mathrm{d}x + \frac{\partial}{\partial y}\,\mathrm{d}y\right)^n f(x, y).$$

**Examples.**

1. Find $d^2z$ for the function $z = \arctan(x/y)$. We have

$$f'_x = \frac{y}{x^2+y^2}, \quad f'_y = -\frac{x}{x^2+y^2}, \quad f''_{xy} = \frac{x^2-y^2}{(x^2+y^2)^2} = f''_{yx},$$

$$f''_{x^2} = -\frac{2xy}{(x^2+y^2)^2}, \quad f''_{y^2} = \frac{2xy}{(x^2+y^2)^2}.$$

Consequently,

$$d^2z = \frac{-2xy\,(dx)^2 + 2(x^2-y^2)\,dx\,dy + 2xy\,(dy)^2}{(x^2+y^2)^2}.$$

2. Find $d^3z$ for the function $z = \sin x \cos y$. We have

$$f'_x = \cos x \cos y, \quad f'_y = -\sin x \sin y, \quad f''_{x^2} = -\sin x \cos y,$$

$$f''_{y^2} = -\sin x \cos y, \quad f''_{xy} = -\cos x \sin y,$$

$$f'''_{x^3} = -\cos x \cos y, \quad f'''_{y^3} = \sin x \sin y,$$

$$f'''_{x^2y} = \sin x \sin y, \quad f'''_{xy^2} = -\cos x \cos y.$$

Consequently,

$$d^3z = -\cos x \cos y\,(dx)^3 + 3 \sin x \sin y\,(dx)^2\,dy$$
$$-3 \cos x \cos y\,dx\,(dy)^2 + \sin x \sin y\,(dy)^3.$$

### 2.12.7. Taylor's Formula for a Function of Two Variables

Like a function of one variable, a function of two variables can be represented as the sum of an $n$th-degree polynomial and a remainder. We shall prove the following theorem.

**Theorem 12.6.** *Let the function $z = f(x, y)$ be continuous together with its partial derivatives up to the order $n+1$ inclusive in a $\delta$-neighbourhood of the point $M(x, y)$. Assume that the point $M_1(x+\Delta x, y+\Delta y)$ belongs to this neighbourhood. Then the increment $\Delta f = f(M_1) - f(M)$ of this function at the point $M$ can be represented as*

$$\Delta f = df(x, y) + \frac{d^2f(x, y)}{2!} + \ldots + \frac{d^nf(x, y)}{n!}$$
$$+ \frac{d^{n+1}f(x+\theta\Delta x, y+\theta\Delta y)}{(n+1)!}, \quad 0 < \theta < 1. \tag{1}$$

Formula (1) is known as *Taylor's formula* for the function $z = f(x, y)$.

*Proof.* We introduce an auxiliary function

$$F(t) = f(x + t\,\Delta x, y + t\,\Delta y),$$

which is a composite function of the independent variable $t$, which varies from 0 to 1, and has an $(n+1)$th derivative with respect to $t$ on the interval $[0, 1]$.

Differentiating the function $F(t)$ with respect to $t$, we obtain

$$F'(t)=f'_x(x+t\,\Delta x,\ y+t\,\Delta y)\,\Delta x+f'_y(x+t\,\Delta x,\ y+t\,\Delta y)\,\Delta y$$
$$=\left(\frac{\partial}{\partial x}\Delta x+\frac{\partial}{\partial y}\Delta y\right)f(x+t\,\Delta x,\ y+t\,\Delta y),$$
$$F''(t)=f''_{xx}(x+t\,\Delta x,\ y+t\,\Delta y)\,(\Delta x)^2+2f''_{xy}(x+t\,\Delta x,\ y+t\,\Delta y)\,\Delta x\,\Delta y$$
$$+f''_{yy}(x+t\,\Delta x,\ y+t\,\Delta y)\,(\Delta y)^2$$
$$=\left(\frac{\partial}{\partial x}\Delta x+\frac{\partial}{\partial y}\Delta y\right)^2 f(x+t\,\Delta x,\ y+t\,\Delta y).$$

By induction we find that

$$F^{(n)}(t)=\left(\frac{\partial}{\partial x}\Delta x+\frac{\partial}{\partial y}\Delta y\right)^n f(x+t\,\Delta x,\ y+t\,\Delta y),$$
$$F^{(n+1)}(t)=\left(\frac{\partial}{\partial x}\Delta x+\frac{\partial}{\partial y}\Delta y\right)^{n+1} f(x+t\,\Delta x,\ y+t\,\Delta y).$$

On the other hand, applying Maclaurin's formula (see 1.6.3, Item 3) to the function $F(t)$ as a function of one variable $t$ and setting $t=1$, we obtain

$$F(1)=F(0)+\frac{F'(0)}{1!}+\frac{F''(0)}{2!}+\ldots+\frac{F^{(n)}(0)}{n!}$$
$$+\frac{F^{(n+1)}(\theta)}{(n+1)!},\quad 0<\theta<1. \tag{2}$$

But

$$F(1)=f(x+\Delta x,\ y+\Delta y)=f(M_1),$$
$$F(0)=f(x,\ y)=f(M),$$
$$F'(0)=\left(\frac{\partial}{\partial x}\Delta x+\frac{\partial}{\partial y}\Delta y\right)f(x,\ y)=\mathrm{d}f(x,\ y),$$
$$F''(0)=\left(\frac{\partial}{\partial x}\Delta x+\frac{\partial}{\partial y}\Delta y\right)^2 f(x,\ y)=\mathrm{d}^2f(x,\ y),$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$F^{(n)}(0)=\left(\frac{\partial}{\partial x}\Delta x+\frac{\partial}{\partial y}\Delta y\right)^n f(x,\ y)=\mathrm{d}^nf(x,\ y),$$
$$F^{(n+1)}(\theta)=\left(\frac{\partial}{\partial x}\Delta x+\frac{\partial}{\partial y}\Delta y\right)^{n+1} f(x+\theta\,\Delta x,\ y+\theta\,\Delta y),$$
$$=\mathrm{d}^{n+1}f(x+\theta\,\Delta x,\ y+\theta\,\Delta y).$$

Taking these relations into account, we have from formula (2)

$$F(1)-F(0)=f(M_1)-f(M)=\Delta f=\mathrm{d}f(x,\ y)+\frac{\mathrm{d}^2f(x,\ y)}{2!}+\ldots+\frac{\mathrm{d}^nf(x,\ y)}{n!}$$
$$+\frac{\mathrm{d}^{n+1}f(x+\theta\,\Delta x,\ y+\theta\,\Delta y)}{(n+1)!},\quad 0<\theta<1,$$

i.e. we have got formula (1). ■

Taylor's formula for a function of two variables resembles Taylor's formula for a function of one variable. But, in reality, if we write in full the expressions for the differentials of the function $f(x, y)$ in formula (1), then we get a more cumbersome and complicated formula than that for a function of one variable.

Taylor's formula for a function of several variables is similar in form.

**Remark.** For $n = 0$ formula (1) yields Lagrange's formula (or a formula of finite increments) for a function of two variables

$$\Delta f = df(x + \theta\,\Delta x,\ y + \theta\,\Delta y)$$
$$= f'_x(x + \theta\,\Delta x,\ y + \theta\,\Delta y)\,\Delta x + f'_y(x + \theta\,\Delta x,\ y + \theta\,\Delta y)\,\Delta y,$$
$$0 < \theta < 1,$$

from which it follows, in particular, that if $f'_x = f'_y \equiv 0$, then the total increment of the function is identically zero and the function $f(x, y)$ is constant.

## 2.12.8. The Extrema of a Function of Two Variables

**1. Definition of an extremum.** Let the function $z = f(x, y)$ be defined in a neighbourhood of the point $M_0(x_0, y_0)$.

**Definition.** *We say that the function $z = f(x, y)$ has a local maximum (minimum) at the point $M_0$ if there is a neighbourhood of the point $M_0$ in which the inequality $f(x, y) \leqslant f(x_0, y_0)$ ($f(x, y) \geqslant f(x_0, y_0)$) holds true for any point $M(x, y)$.*

Points of a local maximum and a local minimum are the *points of extremum*. It follows from the definition that if the function $z = f(x, y)$ has an extremum at the point $M_0$, then the total increment $\Delta z = f(M) - f(M_0)$ of this function at the point $M_0$ satisfies one of the following conditions in a neighbourhood of the point $M_0$:

$$\Delta z \leqslant 0 \text{ (in the case of a local maximum)},$$
$$\Delta z \geqslant 0 \text{ (in the case of a local minimum)}.$$

And, conversely, if one of these inequalities is satisfied in a certain neighbourhood of the point $M_0$, then the function has an extremum at the point $M_0$.

**2. The necessary conditions for an extremum.**

**Theorem 12.7.** *If the function $f(x, y)$ has an extremum at the point $M_0(x_0, y_0)$ and has first-order partial derivatives at the point $M_0$, then, at this point, the first-order partial derivatives are zero, i.e.*

$$f'_x(x_0, y_0) = f'_y(x_0, y_0) = 0. \tag{1}$$

*Proof.* Let us prove, for instance, that the partial derivative $f'_x(x_0, y_0)$ is zero. For that purpose, we shall consider, in the neigh-

bourhood of the point $M_0$, only those points for which $y = y_0$. We get a function $f(x, y_0)$ of one variable $x$, which has an extremum and a derivative $f'_x(x_0, y_0)$ at the point $x = x_0$. Consequently, the necessary condition for an extremum of a function of one variable, $f'_x(x_0, y_0) = 0$, is fulfilled at this point, and this is what we wished to prove.

Similarly, considering the function $f(x_0, y)$ of one variable $y$, we find that $f'_y(x_0, y_0) = 0$. ■

Condition (1) is not a sufficient condition for an extremum. For example, the partial derivatives of the function $z = x^2 - y^2$ are zero at the point (0, 0), but the function has no extremum at this point since it is zero at it and does not retain sign in any neighbourhood of the point (0, 0): if $x = 0$, then $z < 0$, and if $y = 0$, then $z > 0$. The graph of the function $z = x^2 - y^2$ is a hyperbolic paraboloid (see Fig. 160).

Thus condition (1) is only a necessary condition for an extremum. By analogy with a function of one variable, we shall call the points at which it is satisfied the *points of a possible extremum*. Points of this kind are also called *stationary*, or *singular*, *points*.

**3. The sufficient conditions for an extremum.**

**Theorem 12.8.** *Let the function $f(x, y)$ possess continuous second-order partial derivatives at the point $M_0(x_0, y_0)$ of a possible extremum and in a certain neighbourhood of this point. We set*

$$\Delta = \begin{vmatrix} f''_{xx}(x_0, y_0) & f''_{xy}(x_0, y_0) \\ f''_{xy}(x_0, y_0) & f''_{yy}(x_0, y_0) \end{vmatrix}$$

*and then*

(a) *if $\Delta > 0$, then the function has an extremum at the point $M_0$, it being a local maximum for $f''_{xx}(x_0, y_0) < 0$ and a local minimum for $f''_{xx}(x_0, y_0) > 0$;*

(b) *if $\Delta < 0$, then there is no extremum at the point $M_0$.*

*Proof.* (a) Let $\Delta > 0$. We introduce the following designations: $f''_{xx}(x_0, y_0) = A$, $f''_{xy}(x_0, y_0) = B$, $f''_{yy}(x_0, y_0) = C$. By the hypothesis, $f'_x(x_0, y_0) = f'_y(x_0, y_0) = 0$, $A > 0$ (or $A < 0$). According to Taylor's formula (1) from 2.12.7, taken for $n = 1$, the total increment of the function $f(x, y)$ at the point $M_0$ can be written as

$$\Delta f = \frac{1}{2!}[A'(\Delta x)^2 + 2B'\Delta x\Delta y + C'(\Delta y)^2], \tag{1}$$

where $A' = f''_{xx}(x_0 + \theta\Delta x, y_0 + \theta\Delta y)$, $B' = f''_{xy}(x_0 + \theta\Delta x, y_0 + \theta\Delta y)$, $C' = f''_{yy}(x_0 + \theta\Delta x, y_0 + \theta\Delta y)$, $0 < \theta < 1$. From the continuity of the second-order partial derivatives at the point $M_0$ it follows that

$$\lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} A' = f''_{xx}(x_0, y_0) = A > 0 \text{ (or } A < 0),$$

and also

$$\lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} (A'C' - B'^2) = f''_{xx}(x_0, y_0) f''_{yy}(x_0, y_0) - [f''_{xy}(x_0, y_0)]^2 = \Delta > 0.$$

Therefore, for sufficiently small $\Delta y$ and $\Delta x$ we have

$$A' > 0 \text{ (or } A' < 0), \; A'C' - B'^2 = \Delta' > 0.$$

Since $A' \neq 0$, we can write relation (1) in the form

$$\Delta f = \frac{1}{2!} \frac{1}{A'} [A'^2 (\Delta x)^2 + 2A'B' \Delta x \Delta y + A'C' (\Delta y)^2],$$

or, completing it to a perfect square,

$$\Delta f = \frac{1}{2!} \frac{1}{A'} [(A' \Delta x + B' \Delta y)^2 + (A'C' - B'^2) (\Delta y)^2].$$

The expression in brackets is nonnegative and therefore, if $A' > 0$ ($f''_{xx}(x_0, y_0) > 0$), then $\Delta f \geqslant 0$, and, consequently, there is a local minimum at the point $M_0$, and if $A' < 0$ ($f''_{xx}(x_0, y_0) < 0$), then $\Delta f \leqslant 0$, and, consequently, there is a local maximum at the point $M_0$. And this is what we wished to prove.

(b) Let now $\Delta = AC - B^2 < 0$ and, as before, $A = f''_{xx}(x_0, y_0)$, $B = f''_{xy}(x_0, y_0)$, $C = f''_{yy}(x_0, y_0)$. We consider a polynomial

$$A + 2Bx + Cx^2.$$

Since $B^2 - AC > 0$, we can indicate two numbers $x_1$ and $x_2$ such that

$$A + 2Bx_1 + Cx_1^2 > 0, \quad A + 2Bx_2 + Cx_2^2 < 0.$$

We write the total increment of the function $f(x, y)$ at the point $M_0$ in form (1) as we did in Item (a). By virtue of the continuity of the second-order partial derivatives

$$\lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} (A' + 2B'x_1 + C'x_1^2) = A + 2Bx_1 + Cx_1^2 > 0.$$

Consequently, there is a $\delta$-neighbourhood of the point $M_0$ such that if the point $M(x_0 + \Delta x, y_0 + \Delta y)$ belongs to this neighbourhood, then

$$A' + 2B'x_1 + C'x_1^2 > 0. \tag{2}$$

Let us consider now an arbitrary $\delta'$-neighbourhood of the point $M_0$ such that $\delta' \leqslant \delta$. We can choose a number $t > 0$ so small that the point $M_1(x + t, y + tx_1)$ will belong to the $\delta'$-neighbourhood of the point $M_0$. Setting $\Delta x = t$ and $\Delta y = tx_1$ in (1), we get, by virtue of (2), a relation

$$\Delta f = f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0) = \frac{1}{2!} t^2 (A' + 2B'x_1 + C'x_1^2) > 0.$$

As concerns the value $x_2$, reasoning by analogy we find that in the arbitrary $\delta'$-neighbourhood of the point $M_0$ there is a point $M_2(x + \Delta x,\ y + \Delta y)$ for which

$$\Delta f = f(x_0 + \Delta x,\ y_0 + \Delta y) - f(x_0,\ y_0) < 0,$$

i.e. the increment of the function $f(x, y)$ in arbitrarily small neighbourhood of the point $M_0$ does not retain sign and, consequently, there is no extremum at the point $M_0$. ■

**Remark.** If $\Delta = 0$, then the function $f(x, y)$ may have an extremum at the point $M_0$ of a possible extremum and may not possess it.

**Examples.**

1. Test the function $z = x^2 + xy + y^2 - 2x - 3y$ for an extremum. We have

$$f'_x = 2x + y - 2, \quad f'_y = x + 2y - 3.$$

We seek points of a possible extremum. For that purpose, we solve the simultaneous equations

$$\begin{cases} 2x + y - 2 = 0, \\ x + 2y - 3 = 0. \end{cases}$$

Its solutions are $x = 1/3$, $y = 4/3$. Consequently, $M_0(1/3,\ 4/3)$ is a point of a possible extremum. Furthermore, $f''_{xx} = 2$, $f''_{xy} = 1$, $f''_{yy} = 2$ and, consequently, $\Delta = 2 \cdot 2 - 1 = 3$. Since $\Delta = 3 > 0$ and $f''_{xx} = 2 > 0$, the function has a minimum at the point $M_0(1/3,\ 4/3)$.

2. Test the function $z = x^2 - y^2$ for an extremum. We have $f'_x = 2x$, $f'_y = -2y$. Solving the simultaneous equations $2x = 0$, $-2y = 0$, we find that $M_0(0, 0)$ is a point of a possible extremum. Since $f''_{xx} = 2$, $f''_{xy} = 0$, $f''_{yy} = -2$ and, consequently, $\Delta = 2 \cdot (-2) - 0 = -4 < 0$, there is no extremum at the point $M_0(0, 0)$.

3. Test the function $z = x^4 + y^4$ for an extremum. We have $f'_x = 4x^3$, $f'_y = 4y^3$, $f''_{xx} = 12x^2$, $f''_{xy} = 0$, $f''_{yy} = 12y^2$. Solving the simultaneous equations $4x^3 = 0$, $4y^3 = 0$, we find that $M_0(0, 0)$ is a point of a possible extremum. At this point $f''_{xx}(0, 0) = 0$, $f''_{yy}(0, 0) = 0$ and, consequently, $\Delta = 0$. According to the remark made, there may be an extremum at the point $M_0(0, 0)$ or there may be no extremum. In this case there is an extremum since $z > 0$ at all points except for $M_0$ and $z = 0$ at the point $M_0$, i.e. the function has a minimum at the point $M_0$.

4. Test the function $z = x^3 + y^3$ for an extremum. We have $f'_x = 3x^2$, $f'_y = 3y^2$, $f''_{xx} = 6x$, $f''_{xy} = 0$, $f''_{yy} = 6y$. Solving the simultaneous equations $3x^2 = 0$, $3y^2 = 0$, we find that $M_0(0, 0)$ is a point of a possible extremum. At this point $f''_{xx}(0, 0) = 0$, $f''_{yy}(0, 0) = 0$ and, consequently, $\Delta = 0$. In this case there is no extremum at the point $M_0$. Indeed, $z(0, 0) = 0$, $z(x, 0) = x^3$, whence it follows that

$z > 0$ for $x > 0$ and $z < 0$ for $x < 0$, i.e. in any neighbourhood of the point $M_0$ the function has values both larger and smaller than $z(0, 0)$.

### 2.12.9. The Method of Least Squares

In various investigations use is often made of formulas set up on the basis of experiments. One of the best methods of deriving such formulas is the *method of least squares.*

Assume that it is necessary to establish the functional relationship between two variables $x$ and $y$, say, between the temperature and the elongation of a straight metal rod, using the data obtained from the experiment. Using the results of the measurements, we form the following table:

| $x$ | $x_1$ | $x_2$ | ... | $x_i$ | ... | $x_n$ |
|---|---|---|---|---|---|---|
| $y$ | $y_1$ | $y_2$ | ... | $y_i$ | ... | $y_n$ |

Now we find the form of the function $y = f(x)$ proceeding from the positions of the experimental points on the coordinate plane. Assume, for instance, that the points taken from the table are positioned as shown in Fig. 166. In this case it is natural to assume that there is a linear relationship expressed by the formula

$$y = ax + b \qquad (1)$$

between $x$ and $y$. We shall restrict our consideration to a linear relationship.

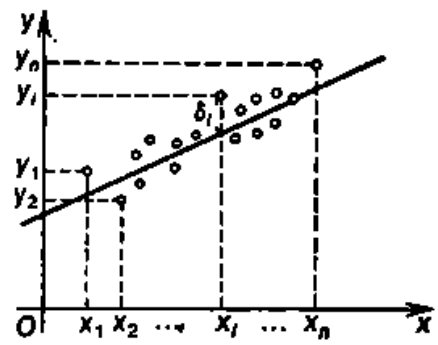


Fig. 166

Since the points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ do not lie exactly on the straight line but only in its vicinity, formula (1) is approximate. Therefore, substituting the values of the coordinates of the points into the expression $y - (ax + b)$, we get relations

$$y_1 - (ax_1 + b) = \delta_1,\ y_2 - (ax_2 + b) = \delta_2,\ \ldots,\ y_n - (ax_n + b) = \delta_n,$$

where $\delta_1, \delta_2, \ldots, \delta_n$ are some numbers which will be called errors.

Then we shall try to choose the coefficients $a$ and $b$ such that the errors will be the least in absolute value. To do this, we use the method of least squares. We consider the sum of the squares of

the errors

$$S(a, b)=\sum_{i=1}^{n}[y_i-(ax_i+b)]^2=\sum_{i=1}^{n}\delta_i^2.$$

Here $x_i$ and $y_i$ are given numbers, and the coefficients $a$ and $b$ are the unknown numbers we have to find proceeding from the condition of the minimum of $S(a, b)$, i.e. we can regard $S(a, b)$ as a function of two variables $a$ and $b$ and test it for an extremum.

Thus the problem has reduced to finding the values of $a$ and $b$ for which the function $S(a, b)$ has a minimum. We have

$$\frac{\partial S}{\partial a}=-2\sum_{i=1}^{n}[y_i-(ax_i+b)]x_i, \quad \frac{\partial S}{\partial b}=-2\sum_{i=1}^{n}[y_i-(ax_i+b)].$$

Equating these partial derivatives to zero, we obtain a linear system of two equations with two unknowns $a$ and $b$:

$$\begin{cases}\sum_{i=1}^{n} y_i x_i=a\sum_{i=1}^{n} x_i^2+b\sum_{i=1}^{n} x_i,\\ \sum_{i=1}^{n} y_i=a\sum_{i=1}^{n} x_i+bn.\end{cases} \tag{2}$$

System (2) is the *normal system of the least squares method*. Using this system, we find the numbers $a$ and $b$ and then, substituting them into formula (1), we get the equation of the required straight line.

The fact that the function $S(a, b)$ has a minimum at the point $M(a, b)$ obtained can be easily established by means of the second-order partial derivatives. We have

$$\frac{\partial^2 S}{\partial a^2}=2\sum_{i=1}^{n} x_i^2, \quad \frac{\partial^2 S}{\partial a\,\partial b}=2\sum_{i=1}^{n} x_i, \quad \frac{\partial^2 S}{\partial b^2}=2n$$

and, consequently,

$$\Delta=\frac{\partial^2 S}{\partial a^2}\frac{\partial^2 S}{\partial b^2}-\left(\frac{\partial^2 S}{\partial a\,\partial b}\right)^2=4n\sum_{i=1}^{n} x_i^2-\left(2\sum_{i=1}^{n} x_i\right)^2.$$

We can write this expression in the form $\Delta=2\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i-x_j)^2$, whence it follows that $\Delta>0$. Since $\frac{\partial^2 S}{\partial a^2}>0$, the function $S(a, b)$ has a minimum at the point $M(a, b)$.

**Example.** Assume that an experiment has yielded five values of the required function $y$ for five values of the argument

| $x$ | $-2$ | 0 | 1 | 2 | 4 |
|---|---|---|---|---|---|
| $y$ | 0.5 | 1 | 1.5 | 2 | 3 |

We seek the functional relationship between $x$ and $y$ in the form of the linear function $y = ax + b$.

To set up the normal system (2) for determining the coefficients $a$ and $b$, we first calculate

$$\sum_{i=1}^{5} y_i x_i = 16.5, \quad \sum_{i=1}^{5} x_i^2 = 25, \quad \sum_{i=1}^{5} x_i = 5, \quad \sum_{i=1}^{5} y_i = 8.$$

System (2) assumes the form

$$\begin{cases} 25a + 5b = 16.5, \\ 5a + 5b = 8. \end{cases}$$

Solving this system, we find that $a = 0.425$ and $b = 1.175$. Consequently, $y = 0.425x + 1.175$ is the equation of the required straight line.

## Chapter 13

## INTEGRATION

In this chapter we shall discuss the main problems connected with the integration of functions of two variables. The definitions and results obtained can be extended to functions of three and more variables.

### 2.13.1. Double Integrals

A double integral is the generalization of the concept of the definite integral to the case of functions of two variables.

**1. Definition and the conditions for the existence of a double integral.** Let $G$ be a bounded closed domain and $z = f(x, y)$ an arbitrary function defined and bounded in this domain.

We assume that the boundary of the domain $G$ consists of a finite number of curves defined by equations of the form $y = f(x)$ or $x = \varphi(y)$, where $f(x)$ and $\varphi(y)$ are continuous functions. A closed polygon, for instance, whose boundary consists of a finite number of seg-

ments which are the graphs of continuous functions of the form $y = kx + b$ or $x = a$, can serve as an example of such a domain. Another example is a domain bounded by an ellipse (the boundary here consists of two curves $y = \pm (b/a)\sqrt{a^2 - x^2}$) etc.

We divide arbitrarily the domain $G$ into $n$ parts $G_i$ which have no interior points in common and whose areas are $\Delta s_i$ ($i = 1, 2, \ldots, n$) (Fig. 167). In each subdomain $G_i$ we choose an arbitrary point $(\xi_i, \eta_i)$ and form a sum

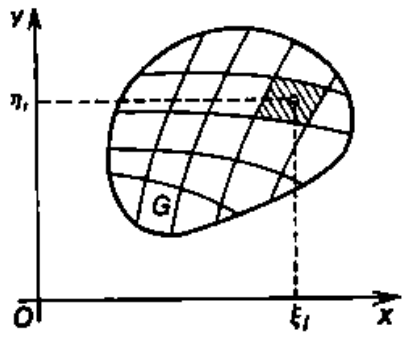


Fig. 167

$$\sigma = \sum_{i=1}^{n} f(\xi_i, \eta_i)\,\Delta s_i, \qquad (1)$$

which is the *integral sum* for the function $f(x, y)$ in the domain $G$. The *diameter* $\mathrm{d}(G)$ of the domain $G$ is the largest distance between the boundary points of the domain. We denote the largest of the diameters of the subdomains $G_i$ by $\lambda$ ($\lambda = \max\limits_{1 \leqslant i \leqslant n} \{\mathrm{d}(G_i)\}$).

**Definition.** *If as $\lambda \to 0$, the integral sum (1) has a limit equal to $I$*, then this limit is called the double integral of the function $f(x, y)$ over the domain $G$ and is denoted by one of the following symbols:*

$$I = \iint\limits_G f(x, y)\,\mathrm{d}s = \iint\limits_G f(x, y)\,\mathrm{d}x\,\mathrm{d}y.$$

In this case the function $f(x, y)$ is said to be *integrable* in the domain $G$, $G$ is called the *domain of integration*, $x$ and $y$ are the *integration variables* and $\mathrm{d}s$ (or $\mathrm{d}x\,\mathrm{d}y$) is the *element of the area*.

When we define a double integral, we assume that the function $f(x, y)$ is bounded. Just as for a function of one variable, this condition is a necessary condition for integrability. It is not sufficient, however, i.e. there are bounded but nonintegrable functions. By way of example, we can consider a function defined on the square $\{(x, y) \mid 0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 1\}$ as follows:

$$f(x, y) = \begin{cases} 1 & \text{if } x \text{ and } y \text{ are rational numbers,} \\ 0 & \text{if } x \text{ or } y \text{ is an irrational number.} \end{cases}$$

The proof of the integrability of this function follows immediately from the definition of a double integral.

To find the sufficient conditions for integrability, it is convenient to use the theory of the Darboux sums, as we did in the case of one

* The number $I$ is the limit of the integral sum (1) as $\lambda \to 0$ if for any $\varepsilon > 0$ there is $\delta > 0$ such that the inequality $|\sigma - I| < \varepsilon$ holds true for $\lambda < \delta$ irrespective of the choice of the points $(\xi_i, \eta_i)$.

variable. This theory can be completely extended to the case of a double integral.*

The following theorem can be proved by analogy with the corresponding theorem for the definite integral.

**Theorem 13.1.** *The function $f(x, y)$ which is continuous in a bounded closed domain $G$ is integrable in this domain.*

We must not assume, however, that a double integral exists only for continuous functions. Here is a more general theorem.

**Theorem 13.2.** *The function $f(x, y)$ bounded in a bounded closed domain $G$ and continuous in it everywhere, except for the points lying on the finite number of curves which are the graphs of continuous functions of the form $y = f(x)$ or $x = \varphi(y)$, is integrable in this domain.*

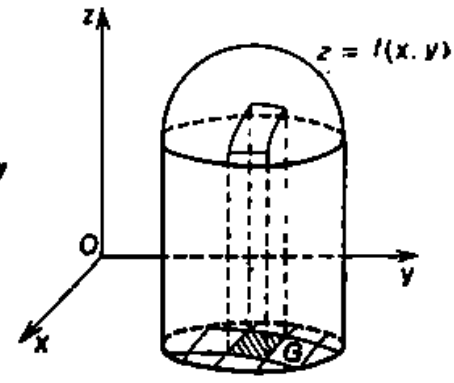


Fig. 168

**2. The geometric meaning of a double integral.** Assume that a body $P$ in space (Fig. 168) is bounded from above by the graph of a continuous and nonnegative function $z = f(x, y)$, which is defined in the domain $G$, from the sides by a cylindrical surface whose directrix is the boundary of the domain $G$ and whose generatrices are parallel to the $z$-axis, and from below by the domain $G$ lying in the $xy$-plane. A body of this shape is a *curvilinear cylinder*.

Just as the problem of calculating the area of a curvilinear trapezoid leads to establishing the geometric meaning of the definite integral, the problem of calculating the volume of the body $P$ leads to the geometric interpretation of the double integral.

Indeed, in this case the integral sum (1) is the sum of the volumes of right cylinders with the base areas $\Delta s_i$ and altitudes $f(\xi_i, \eta_i)$, which can be taken as the approximate value of the volume of the body $P$:

$$v_P \cong \sum_{i=1}^{n} f(\xi_i, \eta_i)\,\Delta s_i.$$

This approximate equality is the more exact, the finer the division of the domain $G$ is. As we pass to the limit when $\lambda \to 0$, the approxima-

* In particular, we can prove that if the function $f(x, y)$ is integrable in the domain $G$, then the limit of both the lower Darboux sums and the upper Darboux sums is equal to $\iint\limits_G f(x, y)\,ds$ as $\lambda \to 0$.

tion becomes exact:

$$v_P = \lim_{\lambda \to 0} \sum_{i=1}^{n} f(\xi_i, \eta_i) \Delta s_i.$$

The function $f(x, y)$ being integrable, the limit of the integral sum exists and is equal to the double integral of this function over the domain $G$. Consequently,

$$v_P = \iint_G f(x, y)\, dx\, dy.$$

Hence follows the geometric meaning of the double integral: *a double integral of a continuous nonnegative function is equal to the volume of a curvilinear cylinder.*

**Remark.** If we set $f(x, y) \equiv 1$ everywhere in the domain $G$, then directly from the definition of the double integral, we get an expression for the area $s$ of the domain $G$ as a double integral:

$$\iint_G 1 \cdot dx\, dy = \lim_{\lambda \to 0} \sum_{i=1}^{n} 1 \cdot \Delta s_i = \lim_{\lambda \to 0} s = s.$$

**3. The properties of the double integral.** The principal properties of the double integral are similar to the corresponding properties of the definite integral. We shall therefore only formulate the properties without proving them.

1°. *If $k$ is an arbitrary number and the function $f(x, y)$ is integrable in the domain $G$, then the function $kf(x, y)$ is also integrable in $G$ and*

$$\iint_G kf(x, y)\, dx\, dy = k \iint_G f(x, y)\, dx\, dy,$$

*i.e. we can put the constant factor before the integral sign.*

2°. *If the functions $f(x, y)$ and $g(x, y)$ are integrable in the domain $G$, then their algebraic sum is also integrable in this domain and*

$$\iint_G [f(x, y) \pm g(x, y)]\, dx\, dy = \iint_G f(x, y)\, dx\, dy \pm \iint_G g(x, y)\, dx\, dy.$$

3°. *If the domain $G$ is the union of the domain $G_1$ and $G_2$ which have no interior points in common and in each of which the function $f(x, y)$ is integrable, then this function is also integrable in the domain $G$ and*

$$\iint_G f(x, y)\, dx\, dy = \iint_{G_1} f(x, y)\, dx\, dy + \iint_{G_2} f(x, y)\, dx\, dy.$$

4°. **The mean value theorem (law of the mean).** *If the function $f(x, y)$ is continuous in the domain $G$, then there is a point $(\xi_i, \eta_i)$ in*

*this domain such that*

$$\iint_G f(x, y)\,dx\,dy = f(\xi_l, \eta_l)\,s,$$

*where s is the area of the figure G.*

We have thus discussed the definition and the principal properties of the double integral, the conditions of its existence and its geometric meaning. We shall discuss now the techniques of calculation of double integrals.

### 2.13.2. Reducing a Double Integral to an Iterated Integral

**1. The case of a rectangular domain.** We shall first consider a double integral over a rectangle $D$ with sides parallel to the coordinate axes.

**Theorem 13.3.** *Assume that in the rectangle* $D = \{(x, y) \mid a \leqslant x \leqslant b,\ c \leqslant y \leqslant d\}$ *there is a double integral*

$$\iint_D f(x, y)\,dx\,dy \tag{1}$$

*for the function* $f(x, y)$. *Assume furthermore that for every* $x$ *from the interval* $[a, b]$ *there is a definite integral*

$$I(x) = \int_c^d f(x, y)\,dy. \tag{2}$$

*Then there is an integral*

$$\int_a^b I(x)\,dx = \int_a^b dx \int_c^d f(x, y)\,dy$$

*(called an iterated integral) and there holds a formula*

$$\iint_D f(x, y)\,dx\,dy = \int_a^b dx \int_c^d f(x, y)\,dy. \tag{3}$$

*Proof.* We use the points $a = x_0 < x_1 < x_2 < \ldots < x_n = b$ and $c = y_0 < y_1 < y_2 < \ldots < y_k = d$ to divide the rectangle $D$ into $nk$ partial rectangles $D_{ij} = \{(x, y) \mid x_{i-1} \leqslant x \leqslant x_i,\ y_{j-1} \leqslant y \leqslant y_j\}$. Then we set $\Delta x_i = x_i - x_{i-1}$, $\Delta y_j = y_j - y_{j-1}$ and designate the greatest lower bound and the least upper bound of the function $f(x, y)$ on the partial rectangle $D_{ij}$ as $m_{ij}$ and $M_{ij}$ respectively (Fig. 169). Then everywhere on the rectangle we have

$$m_{ij} \leqslant f(x, y) \leqslant M_{ij}. \tag{4}$$

In this inequality we set $x = \xi_i$, where $\xi_i$ is an arbitrary point of the subinterval $[x_{i-1}, x_i]$ and then integrate (4) with respect to $y$ in the limits from $y_{j-1}$ to $y_j$. We get

$$m_{ij}\,\Delta y_j \leqslant \int\limits_{y_{j-1}}^{y_j} f(\xi_i,\ y)\,\mathrm{d}y \leqslant M_{ij}\,\Delta y_j. \tag{5}$$

Summing up (5) over all $j$ from 1 to $k$ and using designation (2), we obtain

$$\sum_{j=1}^{k} m_{ij}\,\Delta y_j \leqslant I(\xi_i) \leqslant \sum_{j=1}^{k} M_{ij}\,\Delta y_j. \tag{6}$$

Then, multiplying (6) by $\Delta x_i$ and summing up the result over all $i$ from 1 to $n$, we obtain

$$\sum_{i=1}^{n}\sum_{j=1}^{k} m_{ij}\,\Delta x_i\,\Delta y_j \leqslant \sum_{i=1}^{n} I(\xi_i)\,\Delta x_i \leqslant \sum_{i=1}^{n}\sum_{j=1}^{k} M_{ij}\,\Delta x_i\,\Delta y_j. \tag{7}$$

Let the largest diameter of the partial rectangles $D_{ij}$ tend to zero ($\lambda \to 0$). Then the greatest length $\Delta x_i \to 0$. The extreme terms in (7), which are the lower and upper Darboux sums, tend to the double integral (1) (see the footnote on p. 345). Thus there is a limit of the middle term of (7) as well which is equal to the same double integral. But, by the definition of the definite integral, this limit is

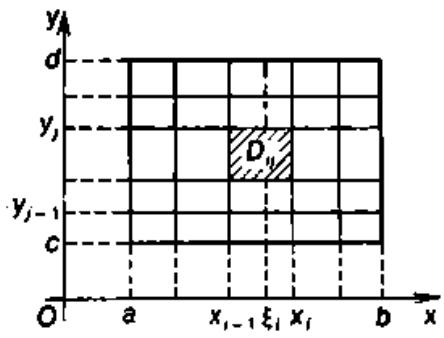


Fig. 169

$$\int\limits_a^b I(x)\,\mathrm{d}x = \int\limits_a^b \mathrm{d}x \int\limits_c^d f(x,\ y)\,\mathrm{d}y.$$

We have thus proved the existence of an iterated integral and formula (3). ■

**Remark.** If we interchange $x$ and $y$ in Theorem 13.3, then we shall prove the existence of the iterated integral

$$\int\limits_c^d I(y)\mathrm{d}y = \int\limits_c^d \mathrm{d}y \int\limits_a^b f(x,\ y)\,\mathrm{d}x$$

and the validity of the formula

$$\iint\limits_D f(x,\ y)\,\mathrm{d}x\,\mathrm{d}y = \int\limits_c^d \mathrm{d}y \int\limits_a^b f(x,\ y)\,\mathrm{d}x. \tag{8}$$

Using formulas (3) and (8), we can reduce the double integral to an iterated integral. For instance, in formula (8) we first integrate with

respect to $x$ for a constant $y$ and then integrate the result with respect to $y$, i.e. successively calculate two definite integrals.

**Example.** Calculate $\iint\limits_D xy\,dx\,dy$, where $D=\{(x, y)\,|\,1\leqslant x\leqslant 2, 1\leqslant y\leqslant 2\}$.

*Solution.* We have

$$\iint\limits_D xy\,dx\,dy=\int\limits_1^2 dy\int\limits_1^2 xy\,dx$$

$$=\int\limits_1^2\left[y\,\frac{x^2}{2}\right]_1^2 dy=\int\limits_1^2\left[2y-\frac{1}{2}\,y\right]dy=\frac{3}{4}\,y^2\Big|_1^2=\frac{9}{4}.$$

**2. The case of a curvilinear domain.**

**Theorem 13.4.** *Let the function $z=f(x, y)$ be defined in the domain $G=\{(x, y)\mid a\leqslant x\leqslant b,\ y_1(x)\leqslant y\leqslant y_2(x)\}$, where $y_1(x)$ and $y_2(x)$ are continuous functions, $y_1(x)\leqslant y_2(x)$ for $a\leqslant x\leqslant b$. Assume also that there is a double integral*

$$\iint\limits_G f(x, y)\,dx\,dy$$

*and for every $x$ from the interval $[a, b]$ there is a definite integral*

$$I(x)=\int\limits_{y_1(x)}^{y_2(x)} f(x, y)\,dy.$$

*Then there is an iterated integral*

$$\int\limits_a^b I(x)\,dx=\int\limits_a^b dx\int\limits_{y_1(x)}^{y_2(x)} f(x, y)\,dy$$

*and the formula*

$$\iint\limits_G f(x, y)\,dx\,dy=\int\limits_a^b dx\int\limits_{y_1(x)}^{y_2(x)} f(x, y)\,dy \tag{9}$$

*holds true.*

*Proof.* We set $c=\min\limits_{[a,b]} y_1(x)$, $d=\max\limits_{[a,b]} y_2(x)$ and place the domain $G$ in a rectangle $D=\{(x, y)\mid a\leqslant x\leqslant b,\ c\leqslant y\leqslant d\}$ (Fig. 170). In this rectangle we consider an auxiliary function

$$F(x, y)=\begin{cases} f(x, y) & \text{at the point of the domain } G,\\ 0 & \text{at the other points of } D.\end{cases}$$

This function satisfies the conditions of the preceding theorem. Indeed, it is integrable in the domain $G$ since it coincides with $f(x, y)$ in it and is integrable in the other part $D - G$ of the rectangle $D$, where it is zero. Consequently, according to Property 3° in 2.13.1, it is also integrable over the whole rectangle $D$, and we have

$$\iint_G F(x, y)\,dx\,dy = \iint_G f(x, y)\,dx\,dy, \quad \iint_{D-G} F(x, y)\,dx\,dy = 0,$$

whence

$$\iint_D F(x, y)\,dx\,dy = \iint_G f(x, y)\,dx\,dy. \tag{10}$$

Furthermore, for every $x$ from $[a, b]$ there is an integral

$$\int_c^d F(x, y)\,dy = \int_c^{y_1(x)} F(x, y)\,dy + \int_{y_1(x)}^{y_2(x)} F(x, y)\,dy + \int_{y_2(x)}^d F(x, y)\,dy$$

since each of the three integrals on the right-hand side exists. Indeed, the intervals $[c, y_1(x)]$ and $[y_2(x), d]$ lie outside the domain $G$ and

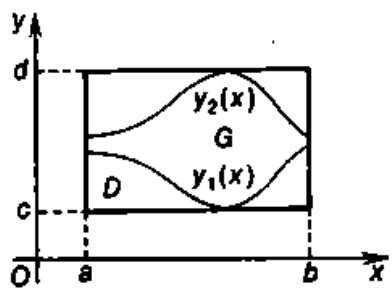



Fig. 170

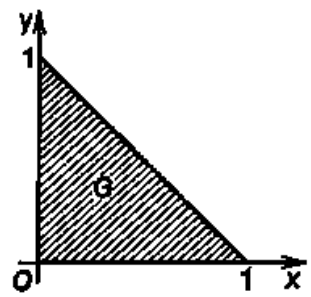



Fig. 171

the function $F(x, y)$ is zero on them. Hence the first and third integrals are zero and the second integral exists by the hypothesis since $F(x, y) = f(x, y)$ on the interval $[y_1(x), y_2(x)]$. Therefore

$$\int_c^d F(x, y)\,dy = \int_{y_1(x)}^{y_2(x)} f(x, y)\,dy. \tag{11}$$

Thus all conditions of Theorem 13.3 are fulfilled for the function $F(x, y)$ and, consequently, the double integral of this function over the rectangle $D$ can be reduced to the iterated integral

$$\iint_D F(x, y)\,dx\,dy = \int_a^b dx \int_c^d F(x, y)\,dy.$$

From this and from formulas (10) and (11) we obtain

$$\iint_G f(x, y)\,dx\,dy = \int_a^b dx \int_{y_1(x)}^{y_2(x)} f(x, y)\,dy,$$

i.e. formula (9). ■

**Remark 1.** If we interchange $x$ and $y$ in Theorem 13.4, then the theorem will assert the existence of the iterated integral

$$\int_c^d I(y)\,dy = \int_c^d dy \int_{x_1(y)}^{x_2(y)} f(x, y)\,dx$$

and the formula

$$\iint_G f(x, y)\,dx\,dy = \int_c^d dy \int_{x_1(y)}^{x_2(y)} f(x, y)\,dx. \tag{12}$$

**Example.** Calculate the integral $\iint_G (x^2+xy+2y^2)\,dx\,dy$ over the domain $G = \{(x, y) \mid 0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 1-x\}$.

*Solution.* The domain $G$ is a triangle bounded by the coordinate axes and the straight line $y = -x+1$ (Fig. 171). Consequently, $y_1(x) = 0$ and $y_2(x) = 1-x$. From formula (9) we have

$$\iint_G (x^2+xy+2y^2)\,dx\,dy = \int_0^1 dx \int_0^{1-x} (x^2+xy+2y^2)\,dy$$

$$= \int_0^1 \left[x^2(1-x) + \frac{x(1-x)^2}{2} + \frac{2(1-x)^3}{3}\right] dx = \frac{7}{24}.$$

We can calculate this integral using formula (2) if we interchange $x$ and $y$ in $G$. Then the triangle is defined by the inequalities $0 \leqslant y \leqslant 1$ and $0 \leqslant x \leqslant 1-y$, whence $x_1(y) = 0$ and $x_2(y) = 1-y$, and it is easy to verify that the integral $\iint_G (x^2+xy+2y^2)\,dx\,dy = \int_0^1 dy \int_0^{1-y} (x^2+xy+2y^2)\,dx$ has the same value.

**Remark 2.** If the domain $G$ does not satisfy the conditions of Theorem 13.4 (e.g. straight lines, vertical or horizontal, cut its boundary at more than two points), then the domain $G$ must be divided into parts each of which would satisfy the conditions of Theorem 13.4, and reduce each of the corresponding double integrals separately to an iterated integral.

### 2.13.3. A Change of Variables in a Double Integral

Let the function $f(x, y)$ be continuous in a bounded closed domain $G$. Then, for the function $f(x, y)$, there is a double integral

$$\iint_G f(x, y)\,dx\,dy. \tag{1}$$

We assume furthermore that, using the formulas

$$x = x(u, v),\ y = y(u, v), \tag{2}$$

we pass to new variables $u$ and $v$. We assume that $u$ and $v$ can be found from (2) in a unique way

$$u = u(x, y), \quad v = v(x, y). \tag{3}$$

Using formulas (3), we put each point $M(x, y)$ from the domain $G$ into correspondence with a point $M^*(u, v)$ on the coordinate plane with rectangular coordinates $u$ and $v$. Let the set of all points $M^*(u, v)$ form a bounded closed domain $G^*$. Formulas (2) are known as *formulas for a transformation of coordinates* and formulas (3) as *formulas for the inverse transformation.*

When we make these assumptions, we can prove that if functions (2) have continuous first-order partial derivatives in the domain $G^*$ and if the determinant

$$\frac{D(x, y)}{D(u, v)} = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} \tag{4}$$

is different from zero in $G^*$, then the formula for a change of variables

$$\iint_G f(x, y)\,dx\,dy = \iint_{G^*} f[x(u, v),\ y(u, v)] \left| \frac{D(x, y)}{D(u, v)} \right| du\,dv \tag{5}$$

holds true for integral (1). Determinant (4) is known as a *functional*, or *Jacobian, determinant* (after the German mathematician Jacobi) of the functions $x = x(u, v)$ and $y = y(u, v)$ with respect to the variables $u$ and $v$.

We can express all we have said in the form of the following theorem.

**Theorem 13.5.** *If transformation* (2) *maps a bounded closed domain $G$ into a bounded closed domain $G^*$ and is one-to-one and if functions* (2) *have continuous first-order partial derivatives and a nonzero Jacobian* (4) *in the domain $G^*$, and the function $f(x, y)$ is continuous in the domain $G$, then the formula for a change of variables* (5) *holds true.*

The proof of the theorem is very complicated and is not given here.*

Both in the double integral and in the definite integral a change of variables is the most important means of reducing it to a form more convenient for calculation.

**Example 1.** Calculate the integral $\iint\limits_G (2x - y)\,dx\,dy$, where $G$ is a parallelogram bounded by the straight lines $x + y = 1$, $x + y = 2$, $2x - y = 1$, $2x - y = 3$ (Fig. 172*a*).

*Solution.* The direct calculation of this integral is very complicated since to reduce it to an iterated integral (first with respect to

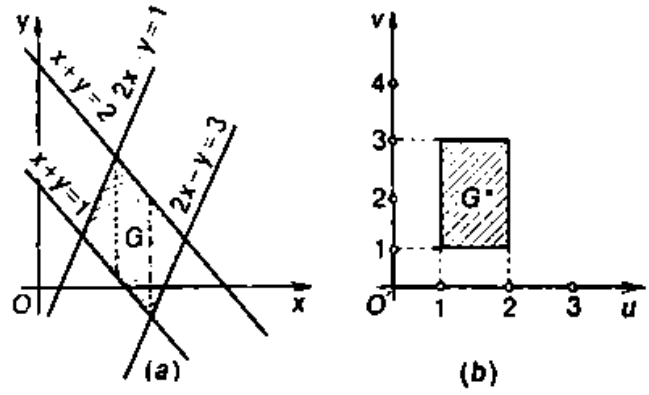



Fig. 172

$y$ and then with respect to $x$), it is necessary to divide the domain $G$ into three subdomains (dash lines in Fig. 172) and then calculate the three corresponding integrals. However, a simple change of variables

$$x + y = u, \quad 2x - y = v \tag{6}$$

makes the solution much simpler. The straight lines $x + y = 1$ and $x + y = 2$ in the system of coordinates $Oxy$ pass into the straight lines $u = 1$ and $u = 2$ in the system of coordinates $O'uv$ (Fig. 172*b*), and the straight lines $2x - y = 1$ and $2x - y = 3$ into the straight lines $v = 1$ and $v = 3$. The parallelogram $G$ is transformed one-to-one into a rectangle $G^*$ which is a simpler domain of integration. It remains to calculate the Jacobian. To do this, we express $x$ and $y$ as $u$ and $v$ from relations (6), i.e. $x = (u + v)/3$ and $y = (2u - v)/3$. Consequently,

$$\frac{D(x, y)}{D(u, v)} = \begin{vmatrix} 1/3 & 1/3 \\ 2/3 & -1/3 \end{vmatrix} = -\frac{1}{9} - \frac{2}{9} = -\frac{1}{3}.$$

* It should also be pointed out that formula (5) holds true in a more general case as well, in particular, Jacobian (4) can vanish at a finite number of points or curves.

From formula (5) we finally obtain

$$\iint\limits_G (2x-y)\,dx\,dy=\iint\limits_{G^*}\frac{1}{3}v\,du\,dv=\frac{1}{3}\int\limits_1^2 du\int\limits_1^3 v\,dv$$

$$=\frac{1}{3}\int\limits_1^2\left[\frac{v^2}{2}\right]_1^3 du=\frac{1}{3}\left(\frac{9}{2}-\frac{1}{2}\right)\int\limits_1^2 du=\frac{1}{3}\cdot 4\cdot 1=\frac{4}{3}.$$

**Remark.** If the integrand function or the equation of the boundary of the domain of integration contains the sum $x^2+y^2$, then we can, in many cases, simplify the integral by transforming it to polar coordinates since in the polar coordinates $x=\rho\cos\varphi$ and $y=\rho\sin\varphi$ the sum assumes a sufficiently simple form $(\rho\cos\varphi)^2+(\rho\sin\varphi)^2=\rho^2$.

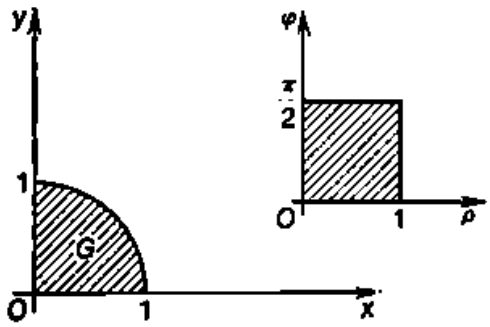


Fig. 173

**Example 2.** Calculate the integral $\iint\limits_G e^{x^2+y^2}\,dx\,dy$, where $G$ is a quarter of the circle $x^2+y^2=1$, which lies in the first quadrant (Fig. 173).

*Solution.* We transform the integral to polar coordinates using the formulas $x=\rho\cos\varphi$ and $y=\rho\sin\varphi$. Then $x^2+y^2=\rho^2$ and

$$\frac{D(x,\ y)}{D(\rho,\ \varphi)}=\begin{vmatrix}\dfrac{\partial x}{\partial\rho} & \dfrac{\partial x}{\partial\varphi}\\ \dfrac{\partial y}{\partial\rho} & \dfrac{\partial y}{\partial\varphi}\end{vmatrix}=\begin{vmatrix}\cos\varphi & -\rho\sin\varphi\\ \sin\varphi & \rho\cos\varphi\end{vmatrix}=\rho(\cos^2\varphi+\sin^2\varphi)=\rho.$$

It can be seen that $\rho$ varies in the domain $G$ in the range from 0 to 1 and $\varphi$ from 0 to $\pi/2$. To put it otherwise, the domain $G$ is transformed into a rectangle $\{(\rho,\ \varphi)\mid 0\leqslant\rho\leqslant 1,\ 0\leqslant\varphi\leqslant\pi/2\}$* (Fig. 173). Thus, using formula (5), we obtain

$$\iint\limits_G e^{x^2+y^2}\,dx\,dy=\int\limits_0^{\pi/2}d\varphi\int\limits_0^1 e^{\rho^2}\rho\,d\rho=\frac{1}{2}(e-1)\int\limits_0^{\pi/2}d\varphi$$

$$=\frac{1}{2}(e-1)\frac{\pi}{2}=\frac{\pi}{4}(e-1).$$

In practical applications it is not always necessary to construct the domain $G^*$. We usually find the range of variation of the new coordinates using the shape of the domain $G$ on the $xy$-plane, as we did at the beginning of this example.

* The Jacobian vanishes for $\rho=0$, but the formula for a change of variables holds good.

### 2.13.4. Some Geometric and Physical Applications of Double Integrals

**1. Calculating the volume.** As is known, the volume $v$ of a curvilinear cylinder, which is bounded from above by the surface $z = f(x, y) > 0$, from below by the plane $z = 0$, and from the lateral

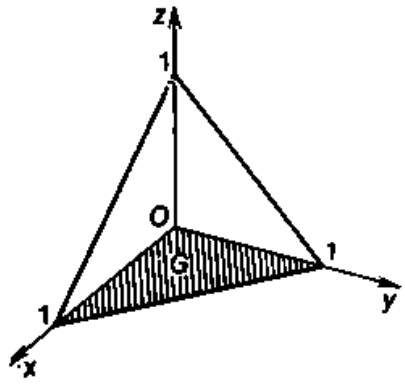



Fig. 174

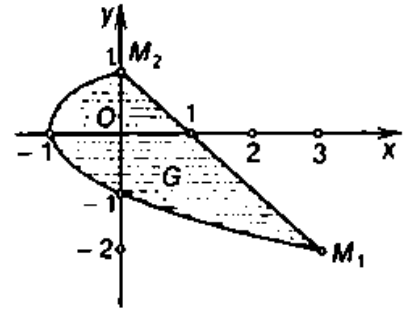



Fig. 175

sides by a cylindrical surface whose generatrices are parallel to the $z$-axis and whose directrix is the contour of the domain $G$, can be found from the formula

$$v = \iint_G f(x, y)\,dx\,dy,$$

i.e. the volume of a body can be calculated by means of double integrals.

**Example 1.** Calculate the volume of a body bounded by the surfaces $x = 0$, $y = 0$, $z = 0$, $x + y + z = 1$ (Fig. 174).

*Solution.* We have

$$v = \iint_G (1 - x - y)\,dx\,dy,$$

where $G$ is a triangular domain of integration bounded by the straight lines $x = 0$, $y = 0$, $x + y = 1$. Writing the limits of integration in the double integral, we obtain

$$v = \int_0^1 dx \int_0^{1-x} (1 - x - y)\,dy = \int_0^1 \left[(1 - x)y - \frac{y^2}{2}\right]_0^{1-x} dx$$

$$= \frac{1}{2}\int_0^1 (1 - x)^2\,dx = \frac{1}{6}.$$

**2. Calculating the area.** We have established (see the remark in 2.13.1) that the area $s$ of the domain $G$ can be calculated by means

of the double integral from the formula

$$s = \iint\limits_G dx\,dy.$$

This formula is more universal than the coorresponding formula which expresses the area of a curvilinear trapezoid with the aid of a definite integral since the given formula is applicable not only to curvilinear trapezoids but also to figures whose position is arbitrary relative to the coordinate axes.

**Example 2.** Calculate the area of the domain $G$ bounded by the curves $y^2 = x + 1$ and $x + y = 1$ (Fig. 175).

*Solution.* The domain $G$ is a figure bounded by the parabola $y^2 = x + 1$ from the left and by the straight line $y = -x + 1$ from the right. Solving simultaneously the equations of the parabola and the straight line, we find their points of intersection $M_1$ (3, $-2$) and $M_2$ (0, 1). Consequently, the required area

$$s = \iint\limits_G dx\,dy = \int\limits_{-2}^{1} dy \int\limits_{y^2-1}^{1-y} dx = \int\limits_{-2}^{1} (2 - y - y^2)\,dy = \frac{9}{2}.$$

When we calculate a double integral using iterated integrals, the setting of the limits of integration is of essential importance. If, in this example, we choose another sequence of the iterated integrals (first with respect to $y$ and then with respect to $x$), then we must first bisect the domain $G$ (by the $y$-axis) since it is bounded from above by a curve given on the intervals $-1 \leqslant x \leqslant 0$ and $0 \leqslant x \leqslant 3$ by two different equations. The result will be naturally the same, but the calculations will be more cumbersome.

It is therefore useful to memorize the following rule: if all the straight lines which are parallel to the $y$-axis enter the domain of integration $G$ across the curve defined by a single equation and leave this domain across the curve defined by a single equation, then it is expedient to take the inner integral with respect to the variable $y$ and the outer integral with respect to the variable $x$; similarly, if all the straight lines which are parallel to the $x$-axis enter the domain of integration across the curve defined by a single equation (across the parabola in this case) and leave this domain across the curve defined by a single equation (across a straight line in this case), then we must take the inner integral with respect to the variable $x$ and the outer integral with respect to the variable $y$, and in this case we may not divide the domain of integration.

**3. Calculating the area of a surface.** Using double integrals, we can calculate the areas of not only plane figures but also of curved surfaces.

Let the surface $S$ be defined by the equation $z = f(x, y)$. The projection of $S$ on the $xy$-plane is a domain $G$ (Fig. 176) and in this do-

main the function $f(x, y)$ is continuous and has continuous partial derivatives $f'_x(x, y)$ and $f'_y(x, y)$. To find the area of the surface $S$, we divide the domain $G$ into $n$ arbitrary parts $G_i$, which have no interior points in common, with areas $\Delta s_i$ $(i = 1, 2, \ldots, n)$ and designate the part of the surface $S$ whose projection on the $xy$-plane is the subdomain $G_i$ as $S_i$. We have thus divided the surface $S$ into $n$ parts.

In each subdomain $G_i$ we choose an arbitrary point $(\xi_i, \eta_i)$ which corresponds to the point $M_i[\xi_i, \eta_i, f(\xi_i, \eta_i)]$ on the surface $S$. Through the point $M_i$ we draw a tangent plane to the surface:

$$f'_x(\xi_i, \eta_i)(x - \xi_i) + f'_y(\xi_i, \eta_i)(y - \eta_i) - (z - z_i) = 0.$$

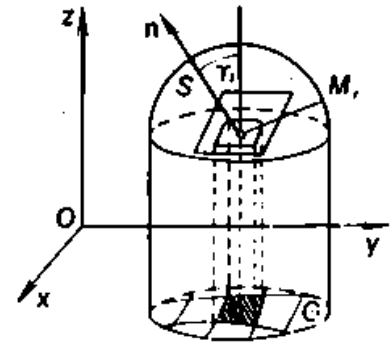


Fig. 176

Here $x, y, z$ are the coordinates of the arbitrary point on the surface and $\xi_i, \eta_i, z_i = f(\xi_i, \eta_i)$ are the coordinates of the point of tangency (see 2.12.4, Item 2). Recall that the vector $\mathbf{n}$ (the normal), which is perpendicular to the tangent plane, has the following coordinates: $\mathbf{n} = \{-f'_x(\xi_i, \eta_i), -f'_y(\xi_i, \eta_i), +1\}$. (Here the vector $\mathbf{n}$ is opposite in direction to the vector $\mathbf{n}$ considered in 2.12.4, Item 2. The vector $\mathbf{n}$ makes an acute angle with the $z$-axis.)

Let us consider the part of the tangent plane whose projection on the $xy$-plane is the subdomain $G_i$. We designate this subdomain as $\sigma_i$ and its area as $\Delta\sigma_i$. We can assume the area $\Delta\sigma_i$ to be approximately equal to the area of the part $S_i$ of the surface, and the sum of all areas of this kind

$$\sum_{i=1}^{n} \Delta\sigma_i$$

to be the appoximate value of the area of the whole surface $S$.

By definition, we assume the limit of such a sum

$$s = \lim_{\lambda \to 0} \sum_{i=1}^{n} \Delta\sigma_i \,^{*} \tag{1}$$

to be the exact value of the area of the surface $S$. Here $\lambda$ is the largest of the diameters of the subdomains $G_i$. We shall prove that this lim-

* The definition of the limit of the sum $\sum_{i=1}^{n} \Delta\sigma_i$ as $\lambda \to 0$ is similar to the definition of the limit of the integral sum for the double integral (see the footnote on p. 344). In what follows, we omit all similar definitions.

it exists and is equal to the double integral

$$s = \iint\limits_G \sqrt{1 + f_x'^2(x, y) + f_y'^2(x, y)}\, dx\, dy. \qquad (2)$$

Next we denote the angle between the vector **n** and the $z$-axis by $\gamma_i$. It is equal to the angle between the tangent plane at the point $M_i$ and the $xy$-plane. Since the subdomain $G_i$ is the projection of $\sigma_i$ on the $xy$-plane, the areas of these subdomains are related as

$$\Delta\sigma_i = \frac{\Delta s_i}{\cos \gamma_i}.$$

Indeed, this formula is known to be valid for triangles. It is evidently valid for plane polygons as well since we can divide a plane polygon into several triangles. It is also valid for any plane figure of area $\Delta\sigma$ bounded by a curve since we can regard its area as the limit of the areas of the polygons inscribed in it.

On the other hand, we know from analytic geometry that

$$\cos \gamma_i = \frac{1}{\sqrt{1 + f_x'^2(\xi_i, \eta_i) + f_y'^2(\xi_i, \eta_i)}}$$

and, consequently,

$$\Delta\sigma_i = \sqrt{1 + f_x'^2(\xi_i, \eta_i) + f_y'^2(\xi_i, \eta_i)}\, \Delta s_i.$$

Substituting the value of $\Delta\sigma_i$ into sum (1), we obtain

$$s = \lim_{\lambda \to 0} \sum_{i=1}^{n} \sqrt{1 + f_x'^2(\xi_i, \eta_i) + f_y'^2(\xi_i, \eta_i)}\, \Delta s_i.$$

The sum under the sign of the limit is the integral sum for the function $\sqrt{1 + f_x'^2(x, y) + f_y'^2(x, y)}$. Since this function is continuous in the domain $G$ by the hypothesis, the limit of this sum exists as $\lambda \to 0$ and is equal to the double integral (2), and this is what we wished to prove.

Relation (2) is a formula which can be used to calculate the area of the surfaces defined by the equation $z = f(x, y)$.

**Example 3.** Calculate the area of the part of the surface $6x + 3y + 2z = 12$ which lies in the first octant (Fig. 177).

*Solution.* Since the function $z = 6 - 3x - (3/2)\, y$ and the domain $G$, which is the projection of this part of the surface on the $xy$-plane, satisfy the conditions formulated above, we can calculate the required area using formula (2). We have

$$f_x'(x, y) = -3, \quad f_y'(x, y) = -\frac{3}{2},$$

$$\sqrt{1 + f_x'^2(x, y) + f_y'^2(x, y)} = \sqrt{1 + 9 + \frac{9}{4}} = \frac{7}{2}.$$

The domain $G$ is a triangle bounded by the $x$-axis and $y$-axis and the straight line $6x + 3y = 12$ obtained from the equation of the given plane for $z = 0$. Writing the limits of integration in the double integral, we obtain

$$s = \int_0^2 dx \int_0^{4-2x} \frac{7}{2} dy = \frac{7}{2} \int_0^2 [y]_0^{4-2x} dx = \frac{7}{2} \int_0^2 (4-2x) dx$$

$$= \frac{7}{2} [4x - x^2]_0^2 = \frac{7}{2} \cdot 4 = 14.$$

**4. Calculating the mass of a plate.** Consider, on the $xy$-plane a material plate, i.e. a domain $G$ over which the mass $m$ is distributed with density $\rho(x, y)$. Proceeding from the density $\rho(x, y)$, we cal-

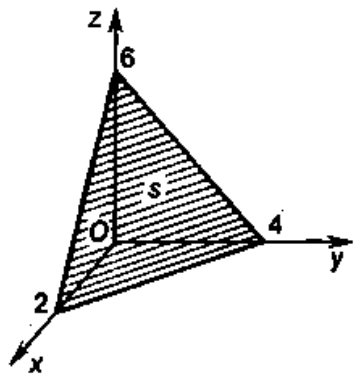



Fig. 177

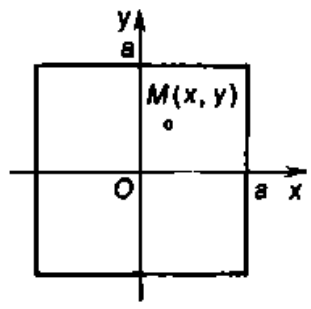



Fig. 178

culate the mass $m$ of the plate assuming that $\rho(x, y)$ is a continuous function. We arbitrarily divide $G$ into $n$ parts $G_i$ $(i = 1, 2, \ldots, n)$ and denote the masses of the subdomains by $m_i$. In each subdomain we arbitrarily take a point $(\xi_i, \eta_i)$. We can consider the mass $m_i$ of each subdomain $G_i$ to be approximately equal to $\rho(\xi_i, \eta_i)\,\Delta s_i$, where $\Delta s_i$ is the area of $G_i$, and the mass $m$ of the whole plate to be approximately equal to the sum

$$m = \sum_{i=1}^{n} m_i \cong \sum_{i=1}^{n} \rho(\xi_i, \eta_i)\,\Delta s_i,$$

which is the integral sum for the continuous function $\rho(x, y)$ in the domain $G$. In the limit, as $\lambda \to 0$, we shall evidently get the exact value of the mass of the plate which is equal to the double integral of the function $\rho(x, y)$ over the domain $G$, i.e.

$$m = \iint_G \rho(x, y)\, dx\, dy. \tag{3}$$

**Example 4.** Find the mass of a square plate with sides $2a$ if the density $\rho(x, y)$ at every point $M(x, y)$ is proportional to the square of

the distance from the point $M$ to the point of intersection of the diagonals and the proportionality factor is $k$.

*Solution.* We choose a system of coordinates as shown in Fig. 178. Then we can find the function $\rho(x, y)$ proceeding from the hypothesis. Let $M(x, y)$ be an arbitrary point of the square plate. Then the square of the distance from the point $M$ to the point of intersection of the diagonals is equal to $x^2 + y^2$. Consequently, the density at the point $M$

$$\rho(M) = \rho(x, y) = k(x^2 + y^2).$$

Using formula (3), we obtain

$$m = \iint\limits_G k(x^2 + y^2)\,dx\,dy.$$

Taking into account that the integrand function is even with respect to $x$ and $y$, and the domain of integration is symmetric about the coordinate axes, we may only calculate the integral over the part of the domain $G$ which lies in the first quadrant, i.e.

$$m = 4k\int_0^a dx\int_0^a (x^2 + y^2)\,dy = 4k\int_0^a \left[x^2y + \frac{y^3}{3}\right]_0^a dx$$

$$= 4k\int_0^a \left(ax^2 + \frac{a^3}{3}\right) dx = 4k\left[\frac{ax^3}{3} + \frac{a^3x}{3}\right]_0^a = 4k\frac{2a^4}{3} = \frac{8}{3}ka^4.$$

**5. Calculating the coordinates of the centre of mass of a plate.** We seek the coordinates of the centre of mass of a plate which occupies a domain $G$ in the $xy$-plane. Let $\rho(x, y)$ be the density of the plate at the point $M(x, y)$ ($\rho(x, y)$ is a continuous function). We divide the domain $G$ into parts $G_i$ ($i = 1, 2, \ldots, n$), choose a point $(\xi_i, \eta_i)$ in each subdomain and consider the mass $m_i$ of each of the parts of the plate to be approximately equal to $\rho(\xi_i, \eta_i)\,\Delta s_i$ ($\Delta s_i$ is the area of $G_i$). If we assume each of the masses to be concentrated at one point, namely, at the point $(\xi_i, \eta_i)$, then for the coordinates $x_c$ and $y_c$ of the centre of mass of such a system of particles we get expressions

$$x_c = \frac{\sum\limits_{i=1}^{n} \xi_i\rho(\xi_i, \eta_i)\,\Delta s_i}{\sum\limits_{i=1}^{n} \rho(\xi_i, \eta_i)\,\Delta s_i}, \quad y_c = \frac{\sum\limits_{i=1}^{n} \eta_i\rho(\xi_i, \eta_i)\,\Delta s_i}{\sum\limits_{i=1}^{n} \rho(\xi_i, \eta_i)\,\Delta s_i}, \tag{4}$$

which are the approximate values of the coordinates of the centre of mass of the plate. To obtain the exact values of the coordinates, it is necessary to pass to the limit in formulas (4) as $\lambda \to 0$. Then the integral sums pass into the corresponding integrals and we find that

the coordinates of the centre of mass of the plate are defined by the formulas

$$x_c = \frac{\iint\limits_G x\rho(x, y)\,dx\,dy}{m}, \quad y_c = \frac{\iint\limits_G y\rho(x, y)\,dx\,dy}{m}, \tag{5}$$

where $m = \iint\limits_G \rho(x, y)\,dx\,dy$ is the mass of the plate.

If the plate is homogeneous, i.e. $\rho = \text{const}$, then the formulas for the coordinates of the centre of mass are simpler:

$$x_c = \frac{\iint\limits_G x\,dx\,dy}{\iint\limits_G dx\,dy}, \quad y_c = \frac{\iint\limits_G y\,dx\,dy}{\iint\limits_G dx\,dy}. \tag{6}$$

The quantities $M_y = \iint\limits_G x\rho(x, y)\,dx\,dy$ and $M_x = \iint\limits_G y\rho(x, y)\,dx\,dy$ in formulas (5) are known as the *static moments* of the plate relative to the $y$-axis and $x$-axis.

Thus the calculation of the coordinates of the centre of mass of the plate reduces to the calculation of three double integrals.

**Example 5.** Find the coordinates of the centre of mass of a homogeneous plate bounded by two parabolas $y^2 = x$ and $x^2 = y$ (Fig. 179).

*Solution.* The coordinates of the centre of mass of the plate can be found from formulas (6). We first calculate the mass of the plate

$$m = \iint\limits_G dx\,dy = \int_0^1 dx \int_{x^2}^{\sqrt{x}} dy = \frac{1}{3}.$$

Then we calculate its static moments relative to the coordinate axes:

$$M_y = \iint\limits_G x\,dx\,dy = \int_0^1 x\,dx \int_{x^2}^{\sqrt{x}} dy = \frac{3}{20},$$

$$M_x = \iint\limits_G y\,dx\,dy = \int_0^1 dx \int_{x^2}^{\sqrt{x}} y\,dy = \frac{3}{20}$$

and then use formulas (6) to find

$$x_c = \frac{M_y}{m} = \frac{3}{20} \div \frac{1}{3} = \frac{9}{20}, \quad y_c = \frac{M_x}{m} = \frac{3}{20} \div \frac{1}{3} = \frac{9}{20}.$$

Thus $x_c = y_c = 9/20$.

**6. Calculating the moment of inertia of a plate.** We know that the moment of inertia of a particle relative to an axis is equal to the

product of the mass of the particle by the square of the distance from the particle to the axis, and the moment of inertia of a system of particles is equal to the sum of the moments of inertia of the particles.

Let the domain $G$ of the $xy$-plane be occupied by a plate with a continuous density $\rho(x, y)$. Dividing the domain $G$ into parts $G_i$ whose areas are $\Delta s_i$ $(i = 1, 2, \ldots, n)$ and choosing a point $(\xi_i, \eta_i)$ in each part, we replace the plate by a system of particles with masses $m_i = \rho(\xi_i, \eta_i)\,\Delta s_i$ and coordinates $(\xi_i, \eta_i)$. The moment of inertia of such a system of point masses, say, relative to the $y$-axis is $\sum_{i=1}^{n} \xi_i^2 \rho(\xi_i, \eta_i)\,\Delta s_i$. We assume this expression to be the approximate value of the moment of inertia of the plate. But it is also the integral sum for the continuous function $x^2\rho(x, y)$. Passing to the limit as $\lambda \to 0$, we get the following formula for the moment of inertia of the plate relative to the $y$-axis:

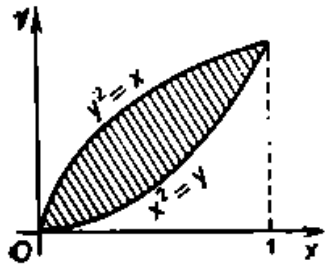


Fig. 179

$$I_y = \iint_G x^2\rho(x, y)\,dx\,dy.$$

Similarly, the moment of inertia of the plate relative to the $x$-axis is

$$I_x = \iint_G y^2\rho(x, y)\,dx\,dy.$$

We shall find the moment of inertia $I_0$ of the plate relative to the origin. Bearing in mind that the moment of inertia of a particle of mass $m$ relative to the origin is $m(x^2 + y^2)$ and reasoning as we did above, we find that

$$I_0 = \iint_G (x^2 + y^2)\,\rho(x, y)\,dx\,dy, \tag{7}$$

i.e.

$$I_0 = I_x + I_y.$$

**Example 6.** Find the moment of inertia of a ring of radius $R$ with constant density $\rho(x, y) = 1$ relative to the origin.

*Solution.* From formula (7) we find that

$$I_0 = \iint_G (x^2 + y^2)\,dx\,dy.$$

We pass to polar coordinates. The equation of a circle (the boundary of the ring) in polar coordinates has the form $\rho = R$ and therefore

$$I_0 = \int_0^{2\pi} d\varphi \int_0^R \rho^2 \rho \, d\rho = \int_0^{2\pi} \left[\frac{\rho^4}{4}\right]_0^R d\varphi = \frac{1}{4}[R^4\varphi]_0^{2\pi} = \frac{\pi R^4}{2}.$$

### 2.13.5. Line Integrals

We shall generalize the concept of the definite integral to the case when the domain of integration is a segment of a curve lying in the plane.

Integrals of this kind are known as *curvilinear*, or *line, integrals*. They have many applications in various divisions of mathematics.

There are two types of line integrals, line integrals of the first kind and those of the second kind.

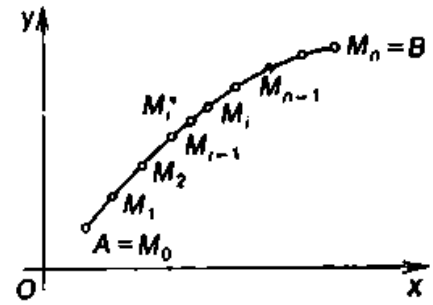



Fig. 180

**1. Line integrals of the first kind.** We consider a smooth or piecewise-smooth* curve $AB$ on the $xy$-plane and assume that the function $z = f(x, y)$ is defined and bounded on the curve $AB$. We arbitrarily divide the curve $AB$ into $n$ parts by the points $A = M_0$, $M_1, M_2, \ldots, M_{i-1}, M_i, \ldots, M_{n-1}, M_n = B$, choose an arbitrary point $M_i^*$ on each of the elementary arcs $M_{i-1}M_i$ (Fig. 180) and form a sum

$$\sum_{i=1}^{n} f(M_i^*)\,\Delta l_i, \tag{1}$$

where $\Delta l_i$ is the length of the arc $M_{i-1}M_i$. The sum (1) is the *integral sum* for the function $z = f(x, y) = f(M)$ over the curve $AB$. We denote the largest of the lengths of the elementary arcs $M_{i-1}M_i$ by $\lambda$ ($\lambda = \max\limits_{1 \leqslant i \leqslant n} \{\Delta l_i\}$).

**Definition.** *If the integral sum* (1) *has a limit equal to $I$ as $\lambda \to 0$, then this limit is a line integral of the first kind of the function $f(x, y)$ over the curve $AB$ and is denoted by one of the following symbols:*

$$I = \int_{AB} f(M)\,dl = \int_{AB} f(x, y)\,dl.$$

* The curve defined by the equations $x = \varphi(t)$ and $y = \psi(t)$, $\alpha \leqslant t \leqslant \beta$, is *smooth* if the functions $\varphi(t)$ and $\psi(t)$ are continuous and have continuous partial derivatives $\varphi'(t)$ and $\psi'(t)$ which do not vanish simultaneously (thus the curve has a tangent at every point). A continuous curve composed of a finite number of smooth pieces is *piecewise smooth*.

In this case the function $f(x, y)$ is said to be *integrable along the curve* $AB$, the curve $AB$ itself is the *path of integration*, $A$ *is* the *initial point* and $B$ the *terminal point of integration.*

It is easy to reduce a line integral of the first kind to a definite integral. Indeed, taking the length of the arc $l$ reckoned from the point $A$ on the curve $AB$ to be a parameter, we get a parametric representation of the curve $x = x(l)$, $y = y(l)$ $(0 \leqslant l \leqslant L)$. In this case the function $f(x, y)$ defined along $AB$ becomes a composite function of the parameter $l$: $f[x(l), y(l)]$. Denoting the value of the parameter $l$ which corresponds to the point $M_i^*$ by $l_i^*$, and the value which corresponds to the point $M_i$ by $l_i$, we rewrite the integral sum (1) as

$$\sum_{i=1}^{n} f[x(l_i^*), y(l_i^*)]\,\Delta l_i, \tag{2}$$

where $\Delta l_i = l_i - l_{i-1}$ and $l_{i-1} \leqslant l_i^* \leqslant l_i$. The sum (2) is the integral sum for the definite integral of the function $f[x(l), y(l)]$ on the interval $[0, L]$. Since the integral sums (1) and (2) are equal, the integrals corresponding to them are equal too, i.e.

$$\int_{AB} f(x, y)\,\mathrm{d}l = \int_0^L f[x(l), y(l)]\,\mathrm{d}l. \tag{3}$$

Note that formula (3) not only expresses the line integral in terms of a definite integral but also proves the existence of a line integral of the function $f(x, y)$ which is continuous along the curve $AB$*.

We have shown that a line integral of the first kind can be immediately reduced to a definite integral, but there is a difference between these integrals, which is as follows. In the integral sum (1) the quantities $\Delta l_i$ are necessarily positive, irrespective of which of the points of the curve $AB$ is initial and which is terminal, i.e.

$$\int_{AB} f(x, y)\,\mathrm{d}l = \int_{BA} f(x, y)\,\mathrm{d}l,$$

whereas the definite integral $\int_a^b f(x)\,\mathrm{d}x$ changes sign when the limits of integration are interchanged. In all other aspects a line integral of the first kind possesses the same properties as a definite integral. This follows directly from formula (3).

---

* The continuity of the function $f(x, y) = f(M)$ along the curve $AB$ means that $\lim\limits_{M \to M_0} f(M) = f(M_0)$ at any point $M_0$ of the curve $AB$, where $M$ is also a point of this curve.

Like the definite integral, the line integral of the first kind has a geometric meaning. Whereas the definite integral $\int_a^b f(x)\,dx$ is the area of a curvilinear trapezoid for $f(x) \geqslant 0$, the line integral $\int_{AB} f(M)\,dl$, for $f(M) \geqslant 0$, is numerically equal to the area of a piece of a cylindrical surface composed of the perpendicular to the $xy$-plane erected at the points $M(x, y)$ of the curve $AB$ and having a variable length $f(M)$ (Fig. 181).

In particular, if $AB$ is not a curve but an interval $[a, b]$ on the $x$-axis, then $f(x, y) = f(x)$, $\Delta l_i = \Delta x_i$, and the line integral is an ordinary definite integral.

Finally, if we set $f(M) \equiv 1$, we get a line integral $\int_{AB} dl$ whose value is the length of the arc of the curve $AB$.

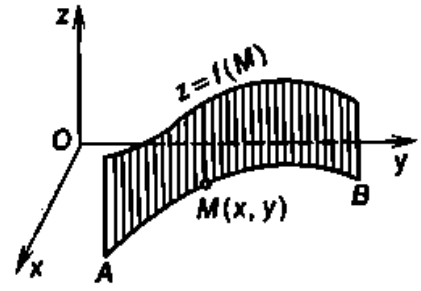


Fig. 181

Thus, using a line integral of the first kind, we can calculate the areas of cylindrical surfaces and the lengths of arcs. In addition, a line integral of the first kind has a variety of applications in physics. As we did in the case of double integrals, we can use it to find the mass of a material curve from its density, moments of inertia relative to the coordinate axes, the coordinates of the centre of mass of such a curve etc.

**2. Calculating line integrals of the first kind.** The calculation of line integrals of the first kind reduces to the calculation of definite integrals.

Let the curve $AB$ be defined by the parametric equations $x = \varphi(t)$, $y = \psi(t)$ $(\alpha \leqslant t \leqslant \beta)$, where $\varphi(t)$ and $\psi(t)$ are functions continuous together with their partial derivatives $\varphi'(t)$ and $\psi'(t)$ and $f(x, y)$ is a function continuous along this curve. We assume, for definiteness, that the point $A$ is associated with the value $t = \alpha$ and the point $B$ with the value $t = \beta$. Then, for any point $M(\varphi(t), \psi(t))$ of the curve $AB$, the length $l$ of the arc $AM$ can be regarded as a function of the parameter $t$: $l = l(t)$ and can be calculated (see 1.8.10, Item 3) using the formula

$$l = l(t) = \int_\alpha^t \sqrt{[\varphi'(t)]^2 + [\psi'(t)]^2}\,dt,$$

whence, according to the rule of differentiation of an integral with respect to the upper limit, we have

$$dl = \sqrt{[\varphi'(t)]^2 + [\psi'(t)]^2}\,dt. \tag{4}$$

Replacing the variable $l = l(t)$ in the definite integral on the right-hand side of formula (3) and taking (4) into account, we obtain

$$\int_{AB} f(x, y)\,dl = \int_0^L f[x(l), y(l)]\,dl$$

$$= \int_\alpha^\beta f[\varphi(t), \psi(t)] \sqrt{[\varphi'(t)]^2 + [\psi'(t)]^2}\,dt. \tag{5}$$

**Example 1.** Calculate the line integral $\int_{AB} y^2\,dl$, where $AB$ is a part of the circle $x = a\cos t$, $y = a\sin t$, $0 \leqslant t \leqslant \pi/2$.

*Solution.* Since

$$y^2 = a^2\sin^2 t, \quad dl = \sqrt{a^2\sin^2 t + a^2\cos^2 t}\,dt = a\,dt,$$

we obtain, from formula (5), a relation

$$\int_{AB} y^2\,dl = \int_0^{\pi/2} a^2\sin^2 t \cdot a\,dt = \frac{a^3}{2}\int_0^{\pi/2} (1 - \cos 2t)\,dt$$

$$= \frac{a^3}{2}\left[t - \frac{\sin 2t}{2}\right]_0^{\pi/2} = \frac{a^3\pi}{4}.$$

In particular, if the curve $AB$ is defined by the equation $y = y(x)$, $a \leqslant x \leqslant b$, where $y(x)$ is a continuously differentiable function *, then we can take $x$ to be a parameter ($t = x$) and obtain, from formula (5), a relation

$$\int_{AB} f(x, y)\,dl = \int_a^b f[x, y(x)] \sqrt{1 + y'^2(x)}\,dx. \tag{6}$$

**Example 2.** Calculate the line integral $\int_{AB} y\,dl$, where $AB$ is the arc of the parabola $y^2 = 2x$ from the point (0, 0) to the point (2, 2).

*Solution.* We have

$$y = \sqrt{2x}, \quad y' = 1/\sqrt{2x}, \quad dl = \sqrt{1 + y'^2}\,dx = \sqrt{1 + \frac{1}{2x}}\,dx.$$

From formula (6) we have

$$\int_{AB} y\,dl = \int_0^2 \sqrt{2x}\sqrt{1 + \frac{1}{2x}}\,dx$$

$$= \int_0^2 \sqrt{2x + 1}\,dx = \frac{1}{3}\left[(2x + 1)^{3/2}\right]_0^2 = \frac{1}{3}(5\sqrt{5} - 1).$$

* I.e. a function which has a continuous derivative.

**Remark.** Formula (4) is of interest by itself. Squaring it, we get $(\mathrm{d}l)^2 = [\varphi'(t)\,\mathrm{d}t]^2 + [\psi'(t)\,\mathrm{d}t]^2 = (\mathrm{d}x)^2 + (\mathrm{d}y)^2$. This relation yields a simple geometric interpretation of the differential of the arc $\mathrm{d}l$. Bearing in mind that the differential of the function $y = y(x)$ is equal to the increment of the ordinate of the tangent (see 1.5.3, Item 1), we find that the differential of the arc $\mathrm{d}l$ (see Fig. 185) is equal to the length of the segment of the tangent to the curve $AB$ from the point of tangency with abscissa $x$ to the point $(x+\mathrm{d}x,\ y+\mathrm{d}y)$, i.e. to the hypotenuse of a right triangle with legs $|\mathrm{d}x|$ and $|\mathrm{d}y|$, and the relation $(\mathrm{d}l)^2 = (\mathrm{d}x)^2 + (\mathrm{d}y)^2$ is the Pythagorean theorem.

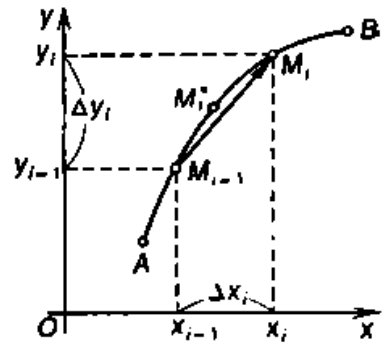


Fig. 182

**3. Line integrals of the second kind.** Assume that two bounded functions $P(x, y)$ and $Q(x, y)$ are defined on the curve $AB$. We divide the curve $AB$ into $n$ parts by the points $A = M_0,\ M_1,\ \ldots,\ M_{i-1},\ M_i,\ \ldots,\ M_n = B$. We designate the projections of the vector $\overrightarrow{M_{i-1}M_i}$ on the coordinate axes as $\Delta x_i$ and $\Delta y_i$ (Fig. 182), take an arbitrary point $M_i^*$ on each elementary arc $M_{i-1}M_i$ and form an integral sum for the function $P(x, y)$ $(Q(x, y))$:

$$\sum_{i=1}^{n} P(M_i^*)\,\Delta x_i \quad \left(\sum_{i=1}^{n} Q(M_i^*)\,\Delta y_i\right). \tag{7}$$

**Definition.** *If the integral sum* (7) *has a limit equal to $I$ as $\lambda \to 0$ ($\lambda = \max\limits_{1 \leqslant i \leqslant n} \{\Delta l_i\}$, $\Delta l_i$ is the length of the arc $M_{i-1}M_i$), then this limit is a line integral of the second kind of the function $P(x, y)$ $(Q(x, y))$ over the curve $AB$ and is designated as*

$$\int\limits_{AB} P(x, y)\,\mathrm{d}x \quad \left(\int\limits_{AB} Q(x, y)\,\mathrm{d}y\right).$$

The sum

$$\int\limits_{AB} P(x, y)\,\mathrm{d}x + \int\limits_{AB} Q(x, y)\,\mathrm{d}y$$

is the *general line integral of the second kind* and is designated as

$$\int\limits_{AB} P(x, y)\,\mathrm{d}x + Q(x, y)\,\mathrm{d}y. \ *$$

* Instead of $P(x, y)$ and $Q(x, y)$ we shall sometimes write simply $P$ and $Q$ and write the line integral as $\int\limits_{AB} P\,\mathrm{d}x + Q\,\mathrm{d}y$.

Just as line integrals of the first kind, line integrals of the second kind can be easily reduced to definite integrals. Indeed, assume that the curve $AB$ is defined by the parametric equations $x = \varphi(t)$, $y = \psi(t)$, $\alpha \leqslant t \leqslant \beta$, where $\varphi(t)$ and $\psi(t)$ are functions continuous together with their partial derivatives $\varphi'(t)$ and $\psi'(t)$, and the point $A$ of the curve is associated with the value $t = \alpha$ and the point $B$ with the value $t = \beta$, and $\varphi'^2(t) + \psi'^2(t) \neq 0$. Let the functions $P(x, y)$ and $Q(x, y)$ be continuous along the curve $AB$. Then there hold formulas

$$\int_{AB} P(x, y)\,dx = \int_{\alpha}^{\beta} P[\varphi(t), \psi(t)]\,\varphi'(t)\,dt,$$

$$\int_{AB} Q(x, y)\,dy = \int_{\alpha}^{\beta} Q[\varphi(t), \psi(t)]\,\psi'(t)\,dt, \tag{8}$$

$$\int_{AB} P(x, y)\,dx + Q(x, y)\,dy$$

$$= \int_{\alpha}^{\beta} \{P[\varphi(t), \psi(t)]\,\varphi'(t) + Q[\varphi(t), \psi(t)]\,\psi'(t)\}\,dt,$$

which reduce the line integrals to definite integrals.

We shall prove the first formula (8):

$$\int_{AB} P(x, y)\,dx = \int_{\alpha}^{\beta} P[\varphi(t), \psi(t)]\,\varphi'(t)\,dt. \tag{9}$$

The second formula can be proved by analogy and the third one results from the summation of the first and the second.

Assume that the points $M_i$ of division of the curve $AB$ are associated with the values $t_i$ of the parameter $t$, the points $M_i^*$ with the values $t_i^*$, i.e. $M_i$ has coordinates $[\varphi(t_i), \psi(t_i)]$ and $M_i^*$ has coordinates $[\varphi(t_i^*), \psi(t_i^*)]$, $i = 1, 2, \ldots, n$. The function $P(x, y)$ on the curve is a composite function of the parameter $t$: $P[\varphi(t), \psi(t)]$. Since the functions $x = \varphi(t)$ and $y = \psi(t)$ are continuous on the interval $[\alpha, \beta]$ and the function $P(x, y)$ is continuous along the curve $AB$, it follows, according to the theorem on the continuity of a composite function, that the function $P[\varphi(t), \psi(t)]$ is continuous on the interval $[\alpha, \beta]$.

Let us form an integral sum (7) for the function $P(x, y)$:

$$\sigma = \sum_{i=1}^{n} P(M_i^*)\,\Delta x_i = \sum_{i=1}^{n} P[\varphi(t_i^*), \psi(t_i^*)]\,\Delta x_i.$$

Since $\Delta x_i = \varphi(t_i) - \varphi(t_{i-1})$, we find from the Newton-Leibniz formula that

$$\Delta x_i = \varphi(t_i) - \varphi(t_{i-1}) = \int_{t_{i-1}}^{t_i} \varphi'(t)\,dt$$

and therefore

$$\sigma = \sum_{i=1}^{n} P[\varphi(t_i^*),\ \psi(t_i^*)] \int_{t_{i-1}}^{t_i} \varphi'(t)\,dt$$

$$= \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} P[\varphi(t_i^*),\ \psi(t_i^*)]\,\varphi'(t)\,dt.$$

On the other hand, since the function $P[\varphi(t), \psi(t)]\varphi'(t)$ is a continuous function on $[\alpha, \beta]$, there is a definite integral for it which appears on the right-hand side of (9). We write it as the sum of integrals over the subinterval $[t_{i-1}, t_i]$:

$$I = \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} P[\varphi(t),\ \psi(t)]\,\varphi'(t)\,dt.$$

Let us consider and estimate the difference

$$\sigma - I = \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \{P[\varphi(t_i^*),\ \psi(t_i^*)] - P[\varphi(t),\ \psi(t)]\}\,\varphi'(t)\,dt. \tag{10}$$

According to Cantor's theorem, the continuity of the function $P[\varphi(t), \psi(t)]$ on $[\alpha, \beta]$ implies its uniform continuity on $[\alpha, \beta]$. And this means that for any $\varepsilon > 0$ there is $\delta > 0$ such that the inequality

$$|P[\varphi(t_i^*), \psi(t_i^*)] - P[\varphi(t), \psi(t)]| < \varepsilon \tag{11}$$

holds true for $\mu = \max\limits_{1 \leqslant i \leqslant n} \{\Delta t_i\} < \delta$. The continuity of the function $\varphi'(t)$ on $[\alpha, \beta]$ implies its boundedness on $[\alpha, \beta]$, i.e. there is a number $k$ such that

$$|\varphi'(t)| \leqslant k. \tag{12}$$

Using (11) and (12), we find the following estimate for difference (10):

$$|\sigma - I| \leqslant \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} |P[\varphi(t_i^*), \psi(t_i^*)] - P[\varphi(t), \psi(t)]| \, |\varphi'(t)| \, dt$$

$$< \varepsilon k \sum_{i=1}^{n} (t_i - t_{i-1}) = \varepsilon k (\beta - \alpha).$$

Since $\varepsilon$ is arbitrary, it follows that

$$\lim_{\mu \to 0} \sigma = I. \tag{13}$$

But as $\lambda = \max\limits_{1 \leqslant i \leqslant n} \{\Delta l_i\} \to 0$, $\mu = \max\limits_{1 \leqslant i \leqslant n} \{\Delta t_i\} \to 0$ $(\Delta t = t_i - t_{i-1})$ as well and vice versa. Indeed, $\Delta l_i = \int_{t_{i-1}}^{t_i} \sqrt{[\varphi'(t)]^2 + [\psi'(t)]^2} \, dt$. The continuity of the functions $\varphi'(t)$ and $\psi'(t)$ on $[\alpha, \beta]$ implies the continuity of the function $\sqrt{[\varphi'(t)]^2 + [\psi'(t)]^2}$ on $[\alpha, \beta]$. But then $m \, \Delta t_i \leqslant \Delta l_i \leqslant M \, \Delta t_i$, where $m$ and $M$ are the minimum and the maximum value of the function $\sqrt{[\varphi'(t)]^2 + [\psi'(t)]^2}$ on $[\alpha, \beta]$ respectively, and $m > 0$ and $M > 0$ by virtue of the condition $\varphi'^2(t) + \psi'^2(t) \neq 0$. It follows from the left-hand inequality that $\mu \to 0$ as $\lambda \to 0$, and from the right-hand inequality it follows that $\lambda \to 0$ as $\mu \to 0$. Consequently, from (13) we have

$$\lim_{\lambda \to 0} \sigma = I,$$

i.e. there is a line integral $\int_{AB} P(x, y) \, dx$ and formula (9) holds true.

The line integral of the second kind possesses properties similar to those of the definite integral, which follows immediately from formulas (8).

As distinct from the line integral of the first kind, the line integral of the second kind depends on the direction of traverse (from $A$ to $B$ or from $B$ to $A$) of the curve $AB$ and changes sign when the direction of traverse of the curve varies, i.e.

$$\int_{AB} P(x, y) \, dx = -\int_{BA} P(x, y) \, dx,$$

$$\int_{AB} Q(x, y) \, dy = -\int_{BA} Q(x, y) \, dy.$$

Indeed, changing the direction in which the curve is traversed, we respectively change the signs of the projections $\Delta x_i$ and $\Delta y_i$ in sums (7) and, consequently, the sums themselves and their limits change signs.

Thus, when calculating line integrals of the second kind, we must take the direction of integration into account.

When $L$ is a closed curve, i.e. when the point $B$ coincides with the point $A$, from the two possible directions of traverse of the closed contour $L$, the direction is termed *positive* when the domain bounded by this contour remains to the left of the point traversing the contour. The opposite direction is termed *negative*.

A line integral taken over a closed contour $L$ traversed in the positive direction is often designated as

$$\oint_L P(x, y)\,dx + Q(x, y)\,dy.$$

**4. Calculating line integrals of the second kind.** Line integrals of the second kind can be calculated by reducing them to definite integrals with the use of formulas (8).

In particular, if the curve $AB$ is defined by an equation of the form $y = y(x)$, $a \leqslant x \leqslant b$, where $y(x)$ is a continuously differentiable function, we can assume $x$ to be a parameter ($t = x$) and obtain, from formulas (8), relations

$$\begin{gathered}\int_{AB} P(x, y)\,dx = \int_a^b P[x, y(x)]\,dx,\\ \int_{AB} Q(x, y)\,dy = \int_a^b Q[x, y(x)]\,y'(x)\,dx,\\ \int_{AB} P(x, y)\,dx + Q(x, y)\,dy\\ = \int_a^b \{P[x, y(x)] + Q[x, y(x)]y'(x)\}\,dx.\end{gathered} \tag{14}$$

We have similar formulas when the curve $AB$ is defined by an equation of the form $x = x(y)$.

**Example 3.** Calculate the integral $\int_{AB} x^2\,dx + xy\,dy$, where $AB$ is a quarter of the circle $x = \cos t$, $y = \sin t$, $0 \leqslant t \leqslant \pi/2$, $A$ corresponds to $t = 0$ and $B$ corresponds to $t = \pi/2$.

*Solution.* We have $x^2 = \cos^2 t$, $dx = -\sin t\, dt$, $xy = \cos t \sin t$, $dy = \cos t\, dt$. From the third formula (8) we find that

$$\int_{AB} x^2\, dx + xy\, dy = \int_0^{\pi/2} (-\cos^2 t \sin t + \cos^2 t \sin t)\, dt = 0.$$

**Example 4.** Calculate the integral $\oint_L (x+y)\, dy$, where $L$ is the contour of a rectangle composed of the straight lines $x = 0$, $y = 0$, $x = 1$, $y = 1$ (Fig. 183).

*Solution.* In Fig. 183 the positive direction of traverse of the contour $L$ is denoted by arrows. Dividing the whole path of integration into parts, we write

$$\oint_L = \int_{AB} + \int_{BC} + \int_{CD} + \int_{DA}.$$

It is easy to note that the integrals taken over the segments $AB$ and $CD$ are zero since $y$ is constant on them and, consequently,

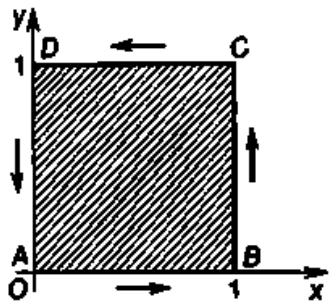



Fig. 183

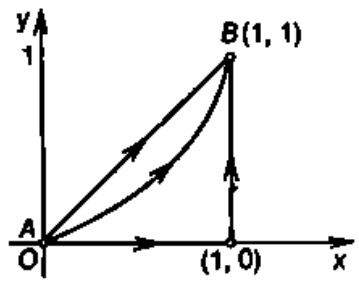



Fig. 184

$dy = 0$. We have therefore to calculate the integrals over the segments $BC$ and $DA$. From a formula similar to the first formula (14) (having replaced $x$ by $y$ and $y(x)$ by $x(y)$) we get

$$\int_{BC} (x+y)\, dy = \int_0^1 (1+y)\, dy = \left[y + \frac{y^2}{2}\right]_0^1 = \frac{3}{2},$$

$$\int_{DA} (x+y)\, dy = \int_1^0 (0+y)\, dy = \left[\frac{y^2}{2}\right]_1^0 = -\frac{1}{2}.$$

Thus we finally have

$$\oint_L (x+y)\, dy = \frac{3}{2} - \frac{1}{2} = 1.$$

**Example 5.** Calculate the integral $\int_{AB} 3x^2 y\, dx + (x^3+1)\, dy$, where

(a) $AB$ is a straight line $y = x$ connecting the points (0, 0) and (1, 1);

(b) $AB$ is a parabola $y = x^2$ connecting the same points;

(c) $AB$ is a polygonal line passing through the points (0, 0), (1, 0), (1, 1) (Fig. 184).

*Solution.* From the third formula (14) we have

$$\text{(a)} \int\limits_{AB} 3x^2y\,dx + (x^3+1)\,dy = \int\limits_0^1 (4x^3+1)\,dx = 2;$$

$$\text{(b)} \int\limits_{AB} 3x^2y\,dx + (x^3+1)\,dy = \int\limits_0^1 (5x^4+2x)\,dx = 2;$$

$$\text{(c)} \int\limits_{AB} 3x^2y\,dx + (x^3+1)\,dy = \int\limits_0^1 3x^2\cdot 0\,dx + \int\limits_0^1 (1+1)\,dy$$

$$= \int\limits_0^1 2dy = 2.$$

Note that taking three different paths connecting the same points, we have got three equal results. This is not accidental. The cause will be disclosed in 2.13.7.

**5. The relations between the line integrals of the first and the second kind.** We designate the angles the directed tangent* to the curve $AB$ at the point $M$ $(x, y)$ makes with the coordinate axes as $\alpha$ and $\beta$ (Fig. 185) and get the following relations:

$$dx = \cos\alpha\,dl, \quad dy = \cos\beta\,dl. \text{ ** } \quad (15)$$

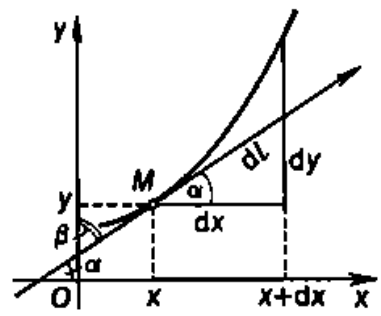


Fig. 185

Replacing $dx$ and $dy$ in the line integrals of the second kind by their expressions (15), we reduce these integrals to line integrals of the first kind:

$$\int\limits_{AB} P(x, y)\,dx = \int\limits_{AB} P(x, y)\cos\alpha\,dl,$$

$$\int\limits_{AB} Q(x, y)\,dy = \int\limits_{AB} Q(x, y)\cos\beta\,dl, \quad (16)$$

$$\int\limits_{AB} P(x, y)\,dx + Q(x, y)\,dy = \int\limits_{AB} [P(x, y)\cos\alpha + Q(x, y)\cos\beta]\,dl.$$

* The positive direction of the tangent will be that which corresponds to the direction of motion of the point along the curve $AB$ from $A$ to $B$.

** See the remark in Item 2.

Thus formulas (16) express line integrals of the second kind in terms of line integrals of the first kind and establish a relationship between them. When the direction of motion of a point along a curve changes to the opposite, $\cos\alpha$, $\cos\beta$, $dx$, $dy$ change sign and formulas (16) remain valid.

It should be pointed out in conclusion that we have discussed line integrals for plane curves. It is, however, easy to extend their definition and properties to space curves.

Assume that $AB$ is a space curve and the functions $f(x, y, z)$, $P(x, y, z)$, $Q(x, y, z)$, $R(x, y, z)$ are defined on this curve. Then, by analogy with the case of a plane curve, we can define a line integral of the first kind $\int\limits_{AB} f(x, y, z)\,dl$ and line integrals of the second kind $\int\limits_{AB} P(x, y, z)\,dx$, $\int\limits_{AB} Q(x, y, z)\,dy$, $\int\limits_{AB} R(x, y, z)\,dz$, $\int\limits_{AB} P(x, y, z)\,dx + Q(x, y, z)\,dy + R(x, y, z)\,dz$. The technique of calculation of such integrals does not differ in essence from the technique of calculation of integrals over a plane curve.

### 2.13.6. Green's Theorem

Green's* theorem establishes a relationship between line integrals and double integrals. It is widely used in the analysis itself and in its applications.

We shall prove this theorem for a closed domain whose boundary is cut by straight lines parallel to the coordinate axes at no more than two points. For simplicity, we shall call domains of this kind *simple domains*. We assume that the contour which bounds the domain is smooth or piecewise smooth.

**Theorem 13.6.** *Assume that $G$ is a simple closed domain bounded by the contour $L$ and let the functions $P(x, y)$ and $Q(x, y)$ be continuous together with their partial derivatives $\frac{\partial P}{\partial y}$ and $\frac{\partial Q}{\partial x}$ in this domain. Then there holds a formula*

$$\iint\limits_G \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\,dy = \oint\limits_L P\,dx + Q\,dy, \tag{1}$$

*which is known as Green's formula.*

*Proof.* Assume that the contour $L$ which bounds the domain $G$ can be defined both by the equations $x = x_1(y)$, $x = x_2(y)$ $(c \leqslant y \leqslant d)$, $x_1(y) \leqslant x_2(y)$, and the equations $y = y_1(x)$,

* Green, George (1793-1841), an English mathematician and physicist.

$y = y_2(x)$ $(a \leqslant x \leqslant b)$, $y_1(x) \leqslant y_2(x)$ (Fig. 186). We shall first consider the domain $G$ defined by the inequalities $a \leqslant x \leqslant b$, $y_1(x) \leqslant y \leqslant y_2(x)$ and reduce the double integral

$$\iint\limits_G \frac{\partial P}{\partial y}\,dx\,dy$$

to a line integral. To do this, we reduce it to an iterated integral and,

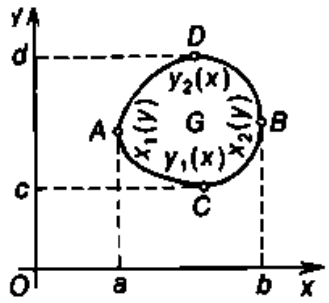



Fig. 186

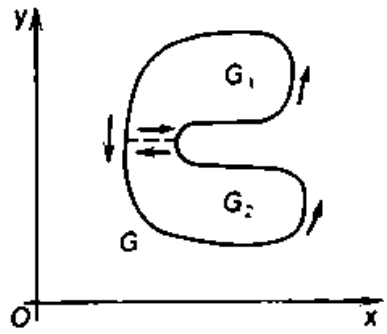



Fig. 187

using the Newton-Leibniz formula, carry out the integration with respect to $y$. We obtain

$$\iint\limits_G \frac{\partial P}{\partial y}\,dx\,dy = \int\limits_a^b dx \int\limits_{y_1(x)}^{y_2(x)} \frac{\partial P}{\partial y}\,dy = \int\limits_a^b [P(x, y_2(x)) - P(x, y_1(x))]\,dx$$

$$= \int\limits_a^b P(x, y_2(x))\,dx - \int\limits_a^b P(x, y_1(x))\,dx.$$

Each of these two definite integrals is equal to the line integral of the second kind taken over a corresponding curve (see formulas (14) in 2.13.5), namely,

$$\int\limits_a^b P(x, y_2(x))\,dx = \int\limits_{ADB} P(x, y)\,dx = -\int\limits_{BDA} P(x, y)\,dx,$$

$$\int\limits_a^b P(x, y_1(x))\,dx = \int\limits_{ACB} P(x, y)dx.$$

Thus we have

$$\iint\limits_G \frac{\partial P}{\partial y}\,dx\,dy = -\Big[\int\limits_{BDA} P(x, y)dx + \int\limits_{ACB} P(x, y)\,dx\Big],$$

i.e.

$$\iint\limits_G \frac{\partial P}{\partial y}\,dx\,dy = -\oint\limits_L P(x, y)dx. \tag{2}$$

By analogy we can prove the formula

$$\iint_G \frac{\partial Q}{\partial x}\,dx\,dy = \oint_L Q(x, y)\,dy \tag{3}$$

(in this case the domain $G$ is defined by the inequalities $c \leqslant y \leqslant d$, $x_1(y) \leqslant x \leqslant x_2(y)$).

Subracting relation (2) from relation (3) term by term, we get the required formula (1). ■

**Remark.** Green's theorem remains true for every closed domain $G$ which can be divided by auxiliary curves into a finite number of simple closed domains. Indeed, let the domain $G$ with boundary $L$ have the form shown in Fig. 187. We divide it into two simple domains $G_1$ and $G_2$ for each of which formula (1) holds true. We write Green's formula separately for $G_1$ and $G_2$ and add the resulting relations together term by term. On the left we have a double integral over the whole domain $G$ and on the right a line integral around the contour $L$ of the domain $G$ since we take the like integral over the auxiliary curve twice in opposite directions and so the integrals cancel out in summation.

**Example.** Using Green's theorem, calculate the line integral $\oint_L (x - y)\,dx + (x + y)\,dy$, where $L$ is a circle $x^2 + y^2 = R^2$.

*Solution.* The functions $P(x, y) = x - y$, $Q(x, y) = x + y$ and $\frac{\partial P}{\partial y} = -1$, $\frac{\partial Q}{\partial x} = 1$ are continuous in the closed circle $x^2 + y^2 = R^2$. Consequently, according to Theorem 13.6, Green's formula is applicable to this integral. We have

$$\oint_L (x-y)dx + (x+y)\,dy = \iint_G [1-(-1)]\,dx\,dy$$

$$= 2\iint_G dx\,dy = 2s = 2\pi R^2.$$

Note that the result obtained can be easily verified by a direct calculation of the given integral.

### 2.13.7. The Conditions for Independence of a Line Integral of the Path of Integration

As we pointed out when solving Example 5 (see 2.13.5, Item 4), in some cases the value of the line integral $\int_{AB} P\,dx + Q\,dy$ does not depend on the path of integration and depends only on the initial and terminal points $A$ and $B$ of the path of integration. Let us

find the conditions under which such an independence exists. Green's theorem plays an important part in the investigation of this problem.

We shall define more precisely the domains which we shall consider in what follows.

**Definition.** *A plane domain G is said to be simply connected if, whatever the closed contour L which lies inside this domain, the part of the plane bounded by this contour entirely belongs to the domain G.*

To make it more picturesque, we can say that a simply connected domain has no "holes". For instance, the interior of a circle, of an ellipse, of a polygon etc is a simply connected domain. The simplest example of a multiply connected domain is a domain included between the circles $x^2 + y^2 = 1$ and $x^2 + y^2 = 3$. Indeed, the circle $x^2 + y^2 = 2$ which lies in this domain contains points in its interior which do not belong to this domain, say, the origin (0, 0).

**Theorem 13.7.** *Let the functions $P(x, y)$ and $Q(x, y)$ be defined and continuous together with their partial derivatives $\frac{\partial P}{\partial y}$ and $\frac{\partial Q}{\partial x}$ in a closed simply connected domain G. Then the following four conditions are equivalent, i.e. the fulfillment of any one of them leads to the fulfillment of the other three conditions:*

(1) *for any closed piecewise smooth curve L which lies in G*

$$\oint_L P\,dx + Q\,dy = 0;$$

(2) *for any two points A and B of the domain G the value of the integral*

$$\int_{AB} P\,dx + Q\,dy$$

*does not depend on the choice of the path of integration which lies entirely in G;*

(3) *the expression $P\,dx + Q\,dy$ is the total differential* of a certain function defined in the domain G. In other words, there is a function $F(x, y)$ defined in G such that*

$$dF = P\,dx + Q\,dy;$$

(4) *everywhere in the domain G we have*

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}. \tag{1}$$

*Proof.* We shall carry out the proof of the theorem using the scheme

$$1 \to 2 \to 3 \to 4 \to 1,$$

* In the theory of line integrals of the second kind the differential of the function $F(x, y)$ is usually called a total differential.

i.e. we shall show that the second condition follows from the first, the third from the second, the fourth from the third, and the fourth condition again leads to the first. We shall thus prove the equivalence of all the conditions.

**1st step:** $1 \to 2$. We consider two arbitrary paths connecting the points $A$ and $B$ in the domain $G$: $ACB$ and $ADB$ are any two piece-

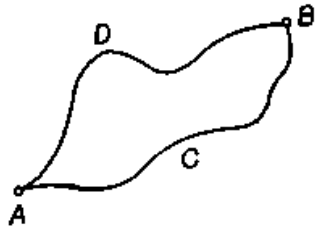



Fig. 188

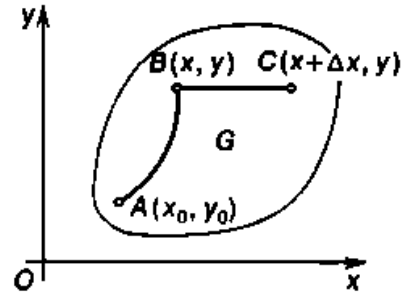



Fig. 189

wise smooth curves (Fig. 188). In sum they form a closed curve $L = ACB + BDA$ which lies in $G$. According to condition (1)

$$\oint_L P\,dx + Q\,dy = 0,$$

but

$$\oint_L P\,dx + Q\,dy = \int_{ACB} P\,dx + Q\,dy + \int_{BDA} P\,dx + Q\,dy$$

$$= \int_{ACB} P\,dx + Q\,dy - \int_{ADB} P\,dx + Q\,dy$$

and, consequently,

$$\int_{ACB} P\,dx + Q\,dy = \int_{ADB} P\,dx + Q\,dy,$$

i.e. condition (2) is fulfilled.

**2nd step:** $2 \to 3$. Let the integral $\int_{AB} P\,dx + Q\,dy$ not depend on the choice of the path of integration but depend only on the points $A$ and $B$. Then, if we fix the point $A$, i.e. $A = A(x_0, y_0)$, then the integral is a function of the coordinates $x$ and $y$ of the point $B = B(x, y)$:

$$\int_{AB} P\,dx + Q\,dy = F(x, y).$$

We shall show that the function $F(x, y)$ is differentiable and that

$$dF = P\,dx + Q\,dy. \tag{2}$$

For that purpose, it is sufficient to prove that there are partial derivatives $\frac{\partial F}{\partial x}$ and $\frac{\partial F}{\partial y}$ at every point $B$ of the domain $G$ and

$$\frac{\partial F}{\partial x} = P(x, y), \quad \frac{\partial F}{\partial y} = Q(x, y). \tag{3}$$

Since $P(x, y)$ and $Q(x, y)$ are continuous in $G$, it follows from (3) that the function $F(x, y)$ is differentiable and relation (2) holds true.

To prove the existence of the partial derivative of the function $F(x, y)$ with respect to $x$ and of the first relation (3), we form a partial increment of the function $F(x, y)$ with respect to $x$ at the point $B(x, y)$

$$\Delta_x F = F(x + \Delta x, y) - F(x, y)$$

$$= \int_{AC} P\,dx + Q\,dy - \int_{AB} P\,dx + Q\,dy = \int_{BC} P\,dx + Q\,dy,$$

where the point $C$ has coordinates $x + \Delta x$ and $y$ (Fig. 189). Since the integral does not depend on the form of the curve by the hypothesis, we assume the path from $B(x, y)$ to $C(x + \Delta x, y)$ to be straight. Then

$$\Delta_x F = \int_{BC} P\,dx + Q\,dy = \int_{BC} P\,dx = \int_{x}^{x+\Delta x} P(x, y)\,dx.$$

Applying the mean value theorem to the last integral, we get

$$\Delta_x F = P(x + \theta\,\Delta x, y)\,\Delta x, \quad 0 < \theta < 1,$$

whence we obtain

$$\frac{\Delta_x F}{\Delta x} = P(x + \theta\,\Delta x, y), \quad 0 < \theta < 1.$$

Consequently,

$$\frac{\partial F}{\partial x} = \lim_{\Delta x \to 0} P(x + \theta\,\Delta x, y) = P(x, y)$$

since $P(x, y)$ is continuous by the hypothesis. We can prove by analogy that $\frac{\partial F}{\partial y} = Q(x, y)$. Thus we have established condition (3).

**3rd step:** $3 \to 4$. Assume that a function $F(x, y)$ defined in the domain $G$ is such that $dF = P\,dx + Q\,dy$. Then we have

$$\frac{\partial F}{\partial x} = P, \quad \frac{\partial F}{\partial y} = Q,$$

and, according to the theorem on the equality of mixed derivatives

$$\frac{\partial Q}{\partial x} = \frac{\partial^2 F}{\partial x\,\partial y} = \frac{\partial^2 F}{\partial y\,\partial x} = \frac{\partial P}{\partial y},$$

i.e. we have obtained the required relation (1).

**4th step:** $4 \to 1$. Assume that condition (4) is fulfilled and let $L$ be a piecewise smooth curve which lies in the domain $G$ and bounds the domain $G^*$. Then, applying Green's theorem to the domain $G^*$ (here we make use of the fact that the domain $G$ is simply connected), we obtain

$$\oint_L P\,dx + Q\,dy = \iint_{G^*} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\,dy.$$

By virtue of condition (4), the integral on the right-hand side is zero and, consequently,

$$\oint_L P\,dx + Q\,dy = 0$$

for every closed contour $L$ which lies in the domain $G$. ■

**Remark.** From the equivalence of conditions (1)-(4) of Theorem 13.7 it follows, in particular, that condition (3) is a necessary and sufficient condition for a line integral to be independent of the choice of the path of integration. But condition (4) is a more convenient necessary and sufficient condition for applications.

Theorem 13.7 makes it easy to decide whether a line integral depends on the path of integration. Thus, for instance, $\int_{AB} e^y\,dx - y\,dy$ depends on the choice of the path in any domain since $\frac{\partial P}{\partial y} = e^y \neq 0 = \frac{\partial Q}{\partial x}$. We must pay attention to the fact that all the conditions of the theorem are essential. Let us consider, for instance, the integral

$$I = \oint_L \frac{-y}{x^2+y^2}\,dx + \frac{x}{x^2+y^2}\,dy,$$

where $L$ is a circle of radius $R$ with centre at the origin. We have

$$P = \frac{-y}{x^2+y^2}, \quad \frac{\partial P}{\partial y} = \frac{-x^2-y^2+2y^2}{(x^2+y^2)^2} = \frac{y^2-x^2}{(x^2+y^2)^2},$$

$$Q = \frac{x}{x^2+y^2}, \quad \frac{\partial Q}{\partial x} = \frac{x^2+y^2-2x^2}{(x^2+y^2)^2} = \frac{y^2-x^2}{(x^2+y^2)^2}.$$

We see that the condition for the independence of the integral of the choice of the path is formally fulfilled, but nevertheless the integral taken around the circle $L$ is not equal to zero. Indeed, defining the circle by the equations $x = R\cos t$ and $y = R\sin t$, we obtain

$$I = \int_0^{2\pi} \frac{-R\sin t(-R\sin t) + R\cos t R\cos t}{R^2\cos^2 t + R^2\sin^2 t}\,dt = \int_0^{2\pi} dt = 2\pi.$$

In actual fact, there is no contradiction with the theorem here. The fact is that one of the conditions of the theorem is not satisfied, i.e. the functions $P$ and $Q$ and their partial derivatives $\frac{\partial P}{\partial y}$ and $\frac{\partial Q}{\partial x}$ are not defined at the point (0, 0) and the circle bounded by circumference $L$, with the point (0, 0) deleted, is no longer a simply connected domain (the origin plays the part of a "hole").

### 2.13.8. Integration of Total Differentials

The consideration of the conditions for the independence of the line integral $\int\limits_{AB} P\,dx + Q\,dy$ of the choice of the path of integration leads to the solution of the problem of integration of total differentials and finding a function proceeding from its total differential.

We have proved that if the functions $P(x, y)$ and $Q(x, y)$ and their partial derivatives $\frac{\partial P}{\partial y}$ and $\frac{\partial Q}{\partial x}$ are continuous in the closed domain $G$, then the expression

$$P\,dx + Q\,dy \tag{1}$$

is the total differential of a function in this domain if and only if

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}.$$

We have proved furthermore that if this equality is satisfied, then the condition

$$dF = P\,dx + Q\,dy$$

is satisfied by the function

$$F(x, y) = \int\limits_{AB} P\,dx + Q\,dy = \int\limits_{(x_0, y_0)}^{(x, y)} P\,dx + Q\,dy. \tag{2}$$

Assume now that expression (1) is the total differential of some function $\Phi(x, y)$. Then $\frac{\partial \Phi}{\partial x} = P$, $\frac{\partial \Phi}{\partial y} = Q$ and the difference $\Phi(x, y) - F(x, y)$ (see the remark to Theorem 12.6) is a constant quantity. Consequently,

$$\Phi(x, y) = F(x, y) + C, \tag{3}$$

where $C$ is a constant. Setting $x = x_0$ and $y = y_0$, we find from (2) that $F(x_0, y_0) = 0$, and from (3) that the value of the constant $C$ is $C = \Phi(x_0, y_0)$. We can now write (3) as

$$F(x, y) = \Phi(x, y) - \Phi(x_0, y_0)$$

and relation (2) as

$$\int_{(x_0,\, y_0)}^{(x,\, y)} P\,dx + Q\,dy = \Phi(x, y) - \Phi(x_0, y_0).$$

If we finally set $x = x_1$ and $y = y_1$, we get a formula

$$\int_{(x_0,\, y_0)}^{(x_1,\, y_1)} P\,dx + Q\,dy = \Phi(x_1, y_1) - \Phi(x_0, y_0) = \Phi(x, y)\Big|_{(x_0,\, y_0)}^{(x_1,\, y_1)}. \qquad (4)$$

Formula (4) is similar to the Newton-Leibniz formula but is valid only when the line integral is independent of the choice of the path of integration.

Using the results obtained, we can now indicate the method of establishing the function $F(x, y)$ whose total differential is the given expression (1).

The formula

$$F(x, y) = \int_{(x_0,\, y_0)}^{(x,\, y)} P\,dx + Q\,dy + C, \qquad (5)$$

where $(x_0, y_0)$ is a fixed point and $C$ is an arbitrary constant, makes it possible to find all functions which have the element of integration as their total differential.

To find $F(x, y)$ using formula (5), it is sufficient to choose any point $(x_0, y_0)$ in the domain $G$ and calculate the line integral over

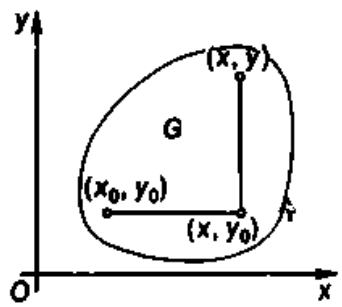



Fig. 190

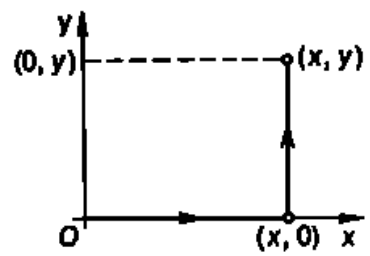



Fig. 191

any curve which connects the points $(x_0, y_0)$ and $(x, y)$. Since the integral in formula (5) does not depend on the choice of the path, it is convenient, for instance, to take the broken line whose segments are parallel to the coordinate axes as the path of integration (Fig. 190). Then

$$\int_{(x_0,\, y_0)}^{(x,\, y)} P\,dx + Q\,dy = \int_{(x_0,\, y_0)}^{(x,\, y_0)} P\,dx + Q\,dy + \int_{(x,\, y_0)}^{(x,\, y)} P\,dx + Q\,dy.$$

Since $y = y_0$ and $dy = 0$ on the interval from $(x_0, y_0)$ to $(x, y_0)$ and $dx = 0$ on the interval from $(x, y_0)$ to $(x, y)$, formula (5) assumes the form

$$F(x, y) = \int_{x_0}^{x} P(x, y_0)\,dx + \int_{y_0}^{y} Q(x, y)\,dy + C,$$

where the first definite integral is calculated for a constant $y$ equal to $y_0$ and the second integral, for a constant $x$.

**Example 1.** Verify whether the expression $(x^2 + 2xy - y^2)\,dx + (x^2 - 2xy - y^2)\,dy$ is the total differential of some function $F(x, y)$ and, if it is, find $F(x, y)$.

*Solution.* In this expression the functions

$$P(x, y) = x^2 + 2xy - y^2,\ Q(x, y) = x^2 - 2xy - y^2 \qquad (6)$$

are continuous together with the partial derivatives $\frac{\partial P}{\partial y} = 2x - 2y$ and $\frac{\partial Q}{\partial x} = 2x - 2y$ which are equal to each other. Consequently, this expression is the total differential of the function $F(x, y)$. To find the function $F(x, y)$, we use relation (2), where $A(x_0, y_0)$ is a fixed point and $B(x, y)$ is a variable point.

In this case it is convenient to take $(0, 0)$ as the point $A(x_0, y_0)$. Taking into account that the line integral $\int_{(0,0)}^{(x,y)} (x^2 + 2xy - y^2)\,dx + (x^2 - 2xy - y^2)\,dy$ does not depend on the path of integration, we choose the path from the point $(0, 0)$ to the point $(x, y)$ in the form of a broken line whose segments are parallel to the coordinate axes. For that purpose, it is sufficient to take the point $(x, 0)$ (or the point $(0, y)$) (Fig. 191). Then one segment of the broken line will lie on the coordinate axis. We have

$$F(x, y) = \int_{(0,0)}^{(x,y)} (x^2 + 2xy - y^2)\,dx + (x^2 - 2xy - y^2)\,dy + C$$

$$= \int_0^x x^2\,dx + \int_0^y (x^2 - 2xy - y^2)\,dy + C = \frac{x^3}{3} + x^2y - xy^2 - \frac{y^3}{3} + C,$$

where $C$ is an arbitrary constant.

In practical applications, to find a function from its total differential, it is convenient to do the following. If

$$\frac{\partial F}{\partial x} = P,\ \frac{\partial F}{\partial y} = Q,$$

then, integrating the first relation with respect to $x$, we get

$$F(x, y) = \int P\,dx + f_1(y), \tag{7}$$

and integrating the second relation with respect to $y$, we get

$$F(x, y) = \int Q\,dy + f_2(x), \tag{8}$$

where $f_1(y)$ and $f_2(x)$ are arbitrary functions. If we choose the functions $f_1(y)$ and $f_2(x)$ such that the right-hand sides of relations (7) and (8) coincide, the resulting function $F(x, y)$ will be the function whose total differential coincides with the expression $P\,dx + Q\,dy$.

Thus, for instance, let $dF = (2xy + 1)\,dx + (x^2 + 3y^2)\,dy$. Integrating the coefficient in $dx$ with respect to $x$, we obtain

$$\int (2xy + 1)\,dx = x^2y + x + f_1(y), \tag{9}$$

and integrating the coefficient in $dy$ with respect to $y$, we get

$$\int (x^2 + 3y^2)\,dy = yx^2 + y^3 + f_2(x). \tag{10}$$

The right-hand sides of relations (9) and (10) coincide if we set $f_1(y) = y^3 + C$ and $f_2(x) = x + C$.

Thus we have

$$F(x, y) = yx^2 + y^3 + x + C.$$

**Example 2.** Calculate the line integral $\int\limits_{(-1,\,2)}^{(2,\,3)} y\,dx + x\,dy$.

*Solution.* In this case the functions

$$P = y,\ Q = x,\ \frac{\partial P}{\partial y} = 1,\ \frac{\partial Q}{\partial x} = 1$$

are continuous and the partial derivatives are equal to each other. This means that the expression $y\,dx + x\,dy$ is the total differential $dF(x, y)$ and the given integral does not depend on the path of integration. Using relations (7) and (8), we find that $F(x, y) = xy$, and from formula (4) we obtain

$$\int\limits_{(-1,\,2)}^{(2,\,3)} y\,dx + x\,dy = xy\,\Big|_{(-1,\,2)}^{(2,\,3)} = 2\cdot 3 - (-1)\cdot 2 = 6 + 2 = 8.$$

Note that we can calculate the given integral directly if we take, as the path of integration, say, a broken line connecting the points $(-1, 2)$, $(2, 2)$, $(2, 3)$ whose segments are parallel to the coordinate axes (do it yourself!).

### 2.13.9. Some Applications of the Second-Kind Line Integrals

Like integrals of the first kind, line integrals of the second kind have a variety of applications in geometry, physics and engineering. We shall consider here two problems, the problem of calculating the areas of plane figures and that of finding the work done by a force.

**1. Using Green's formula to calculate the area.** Assume that $G$ is a domain with a boundary $L$ and $s$ is the area of this domain. It is known that for $f(x, y) \equiv 1$ the double integral $\iint\limits_G f(x, y)\,dx\,dy$ expresses the area of the domain $G$. Therefore, if in Green's formula we choose the functions $P(x, y)$ and $Q(x, y)$ such that $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \equiv 1$, then the area $s$ of the domain $G$ can be found from the formula

$$s = \iint\limits_G dx\,dy = \oint\limits_L P\,dx + Q\,dy.$$

We set $Q(x, y) = x$ and $P(x, y) = 0$, then $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = 1$ and

$$s = \oint\limits_L x\,dy.$$

Setting $P(x, y) = -y$ and $Q(x, y) = 0$, we find by analogy that

$$s = -\oint\limits_L y\,dx,$$

and for $P(x, y) = -y/2$ and $Q(x, y) = x/2$, we have

$$s = \frac{1}{2}\oint\limits_L x\,dy - y\,dx. \tag{1}$$

We have thus obtained three formulas to calculate the areas of plane figures bounded by the contour $L$.

**Example 1.** Find the area of the figure bounded by an ellipse.

*Solution.* Let us calculate the area using formula (1), for instance. From the parametric equations of the ellipse $x = a\cos t$ and $y = b\sin t$, $0 \leqslant t \leqslant 2\pi$, we have $dx = -a\sin t\,dt$ and $dy = b\cos t\,dt$, and from formula (1) we obtain

$$s = \frac{1}{2}\oint\limits_L x\,dy - y\,dx$$

$$= \frac{1}{2}\int\limits_0^{2\pi} (a\cos t\, b\cos t + b\sin t\, a\sin t)\,dt = \frac{ab}{2}\int\limits_0^{2\pi} dt = \pi ab.$$

**2. The work done by a force.** It is known that the work done by a variable force $\mathbf{F}(x)$ directed along the $x$-axis to displace a particle along the $x$-axis from the point $x = a$ to the point $x = b$ $(a < b)$ can be found by means of a definite integral using the formula $A = \int_a^b F(x)\,dx$ (see 1.6.8, Item 6). Let us consider a more general problem.

Assume that under the action of the force $\mathbf{F}$ a particle moves along a continuous plane curve $BC$ in the direction from $B$ to $C$. We assume the force $\mathbf{F}$ to be variable and dependent on the position of the

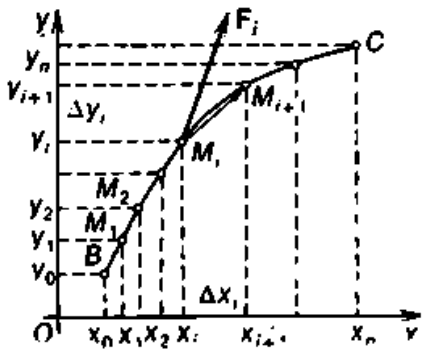



Fig. 192

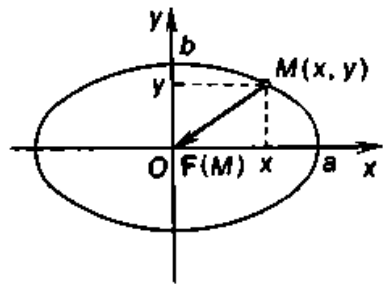



Fig. 193

particle on the curve $BC$. We calculate the work done by the force $\mathbf{F}$ displacing the particle from $B$ to $C$. To do this, we arbitrarily divide the curve $BC$ into $n$ parts by the points $B = M_0, M_1, M_2, \ldots, M_i, M_{i+1}, \ldots, M_n = C$ (Fig. 192). On the interval $M_iM_{i+1}$ we approximate the force $\mathbf{F}$ by a constant value equal to its value at the point $M_i$, i.e. $\mathbf{F}_i \cong \mathbf{F}(M_i)$, and the movement of the particle along the arc $M_iM_{i+1}$ by the movement along the interval $M_iM_{i+1}$. Then we can take the work done by the constant force $\mathbf{F}(M_i)$ along the interval $M_iM_{i+1}$ as the approximate value of the work $A_i$ of the variable force $\mathbf{F}$ along the arc $M_iM_{i+1}$, i.e.

$$A_i \cong \mathbf{F}(M_i)\cdot\overrightarrow{M_iM_{i+1}}.$$

The right-hand side of this approximate equality is the scalar product of two vectors $\mathbf{F}(M_i)$ and $\overrightarrow{M_iM_{i+1}}$. It is equal to the sum of the products of the corresponding coordinates of these vectors, i.e. if $\mathbf{F}(M_i) = \{P(M_i),\ Q(M_i)\}$ and $\overrightarrow{M_iM_{i+1}} = \{\Delta x_i,\ \Delta y_i\}$, then $A_i \cong P(M_i)\,\Delta x_i + Q(M_i)\,\Delta y_i$. Summing up $A_i$ over all the values of $i$ from 1 to $n$, we get the approximate value of the work $A$ along

the whole curve $BC$:

$$A \cong \sum_{i=1}^{n} [P(M_i)\,\Delta x_i + Q(M_i)\Delta y_i]$$

$$= \sum_{i=1}^{n} P(M_i)\,\Delta x_i + \sum_{i=1}^{n} Q(M_i)\,\Delta y_i. \qquad (2)$$

For the exact value of the work $A$ we take the limit to which its approximate value tends as the largest of the lengths of the arcs $M_iM_{i+1}$ tends to zero. But, on the other hand, sum (2) is the sum of two integral sums for the functions $P(x, y)$ and $Q(x, y)$ defined on the curve $BC$. By definition, the limit of this sum is a line integral of the second kind. Consequently, the work done by the force can be found from the formula

$$A = \int_{BC} P\,dx + Q\,dy, \qquad (3)$$

where $P$ and $Q$ are the coordinates (or the projections on the coordinate axes) of the force $\mathbf{F}$.

If we consider this problem in space rather than on the plane, then its solution reduces to the calculation of a line integral of the second kind over a space curve by the formula

$$A = \int_{BC} P\,dx + Q\,dy + R\,dz.$$

**Example 2.** Calculate the work done by the force $\mathbf{F}(x, y)$ displacing a particle along an ellipse in the positive direction if at every point $(x, y)$ of the ellipse the force is directed towards the centre of the ellipse and is equal in value to the distance from the point $(x, y)$ to the centre of the ellipse (Fig. 193).

*Solution.* By the hypothesis, $|\mathbf{F}(x, y)| = \sqrt{x^2 + y^2}$; the coordinates of the force $\mathbf{F}(x, y)$ are $P = -x$, $Q = -y$ (the minus sign shows that the force is directed to the point (0, 0)). Using formula (3), we find that $A = -\oint_L x\,dx + y\,dy$, where $L$ is the ellipse $x = a\cos t$, $y = b\sin t$, $0 \leqslant t \leqslant 2\pi$. Consequently,

$$A = \frac{a^2 - b^2}{2} \int_0^{2\pi} \sin 2t\,dt = \frac{a^2 - b^2}{4}(-\cos 2t)\Big|_0^{2\pi} = 0.$$

Note that from the equality of the integral to zero it follows that the element of integration is the total differential of some function (find the function yourself!).

### 2.13.10. Triple Integrals

At the beginning of the chapter we introduced the concept of a double integral of a function of two variables. We shall define now an integral of a function of three variables, the so-called triple integral. Triple integrals, as well as double integrals, have many applications in various problems of physics and geometry.

**1. Definition of a triple integral.** A triple integral is an analogue of a double integral and is introduced for a function of three variables.

Let a bounded function $f(M) = f(x, y, z)$ be defined in a bounded closed domain $V$ of a three-dimensional space. We divide the domain $V$ into $n$ arbitrary subdomains, which have no interior points in common, with volumes $\Delta v_1, \Delta v_2, \ldots, \Delta v_n$. We take an arbitrary point $M_i(\xi_i, \eta_i, \zeta_i)$ in each subdomain and form a sum

$$\sum_{i=1}^{n} f(\xi_i, \eta_i, \zeta_i)\,\Delta v_i, \tag{1}$$

which is known as the *integral sum* for the function $f(x, y, z)$ over the domain $V$. We denote the largest of the diameters of the subdomains by $\lambda$.

**Definition.** *If the integral sum* (1) *has a limit equal to $I$ as $\lambda \to 0$, then this limit is the triple integral of the function $f(x, y, z)$ over the domain $V$ and is denoted by one of the following symbols*:

$$I = \iiint\limits_V f(x, y, z)\,dv = \iiint\limits_V f(x, y, z)\,dx\,dy\,dz.$$

In this case the function $f(x, y, z)$ is said to be *integrable in the domain* $V$, $V$ is the *integration domain*, $x, y, z$ are the *integration variables*, and $dv$ (or $dx\,dy\,dz$) is the *element of volume*.

Since the results obtained for double integrals, together with their proofs, can be extended to triple integrals, we shall only formulate the statements and give brief explanations.

Triple integrals are direct generalizations of double integrals to the case of a three-dimensional space. The necessary and sufficient conditions for their existence and their properties are similar to those of double integrals. If we set $f(x, y, z) \equiv 1$ everywhere in the domain $V$, then we can derive formula for calculating the volume of the body in the domain $V$ using the definition of a triple integral:

$$v = \iiint\limits_V dv = \iiint\limits_V dx\,dy\,dz.$$

**2. Calculating triple integrals.** As in the case of double integrals, the calculation of triple integrals reduces to the calculation of integrals of a lower multiplicity.

Let us consider a domain $V$ bounded by the surfaces $z = z_1(x, y)$ and $z = z_2(x, y)$ from below and from above and by a cylindrical surface from the lateral sides and assume that the domain $G$ is the projection of the domain $V$ on the $xy$-plane (Fig. 194) in which the functions $z_1(x, y)$ and $z_2(x, y)$ are defined and continuous. We assume furthermore that every straight line which is parallel to the

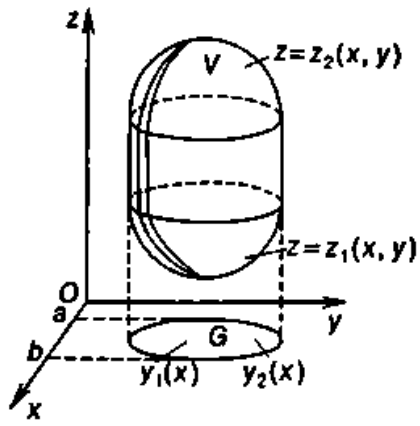



Fig. 194

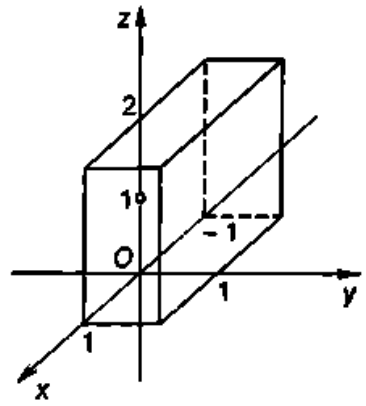



Fig. 195

$z$-axis cuts the boundary of the domain $V$ at no more than two points. Then, for any function $f(x, y, z)$ continuous in the domain $V$, there holds a formula

$$\iiint\limits_V f(x, y, z)\,dx\,dy\,dz = \iint\limits_G dx\,dy \int\limits_{z_1(x,\,y)}^{z_2(x,\,y)} f(x, y, z)\,dz,$$

which makes it possible to reduce the calculation of a triple integral to a successive calculation of an inner definite integral with respect to the variable $z$ (for constant $x$ and $y$) and of an outer double integral over the domain $G$.

The expression

$$I(x, y) = \int\limits_{z_1(x,\,y)}^{z_2(x,\,y)} f(x, y, z)\,dz$$

is a function of two variables. If the conditions of Theorem 13.4 are fulfilled for this function and the domain $G$ over which it is integrated, then, passing from the double integral $\iint\limits_G I(x, y)\,dx\,dy$ to an iterated integral, we get a formula

$$\iiint\limits_V f(x, y, z)\,dx\,dy\,dz = \int\limits_a^b dx \int\limits_{y_1(x)}^{y_2(x)} dy \int\limits_{z_1(x,\,y)}^{z_2(x,\,y)} f(x, y, z)\,dz, \qquad (2)$$

which reduces the calculation of a triple integral to the successive calculation of three definite integrals. The order of integration may vary, i.e. we can interchange the variables $x$, $y$, $z$ in formula (2).

In particular, if $V$ is a parallelepiped with faces $x = a$, $x = b$ $(a < b)$, $y = c$, $y = d$ $(c < d)$, $z = k$, $z = l$ $(k < l)$, then formula (2) assumes the form

$$\iiint\limits_V f(x, y, z)\,dx\,dy\,dz = \int_a^b dx \int_c^d dy \int_k^l f(x, y, z)\,dz. \qquad (3)$$

In this case we can carry out the integration in any order.

**Example 1.** Calculate the integral $\iiint\limits_V (x + y - z)\,dx\,dy\,dz$, where $V$ is a parallelepiped bounded by the planes $x = -1$, $x = +1$, $y = 0$, $y = 1$, $z = 0$, $z = 2$ (Fig. 195).

*Solution.* From formula (3) we have

$$\iiint\limits_V (x+y-z)\,dx\,dy\,dz = \int_{-1}^{1} dx \int_0^1 dy \int_0^2 (x+y-z)\,dz$$

$$= \int_{-1}^{1} dx \int_0^1 \left[xz+yz-\frac{z^2}{2}\right]_0^2 dy = \int_{-1}^{1} dx \int_0^1 (2x+2y-2)\,dy$$

$$= \int_{-1}^{1} [2xy+y^2-2y]_0^1\,dx = \int_{-1}^{1} (2x-1)\,dx = [x^2-x]_{-1}^{1} = -2.$$

**Example 2.** Calculate the integral $\iiint\limits_V (x + y + z)\,dx\,dy\,dz$, where $V$ is a pyramid bounded by the plane $x + y + z = 1$ and the coordinate planes $x = 0$, $y = 0$, $z = 0$ (Fig. 196).

*Solution.* The domain $V$ is projected onto the $xy$-plane into the triangle $G$ bounded by the straight lines $x = 0$, $y = 0$, $y = 1 - x$. From formula (2) we have

$$\iiint\limits_V (x+y+z)\,dx\,dy\,dz$$

$$= \int_0^1 dx \int_0^{1-x} dy \int_0^{1-x-y} (x+y+z)\,dz$$

$$= \int_0^1 dx \int_0^{1-x} \left[xz+yz+\frac{z^2}{2}\right]_0^{1-x-y} dy$$

$$= \frac{1}{2} \int_0^1 \left[ y - yx^2 - xy^2 - \frac{y^3}{3} \right]_0^{1-x} dx$$

$$= \frac{1}{6} \int_0^1 (2 - 3x + x^3)\, dx = \frac{1}{6} \left[ 2x - \frac{3x^2}{2} + \frac{x^4}{4} \right]_0^1 = \frac{1}{6} \cdot \frac{3}{4} = \frac{1}{8}.$$

**3. A change of variables in a triple integral.** For triple integrals, as well as for double integrals, there hold formulas for the transition

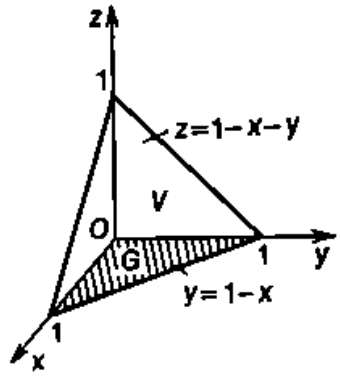



Fig. 196

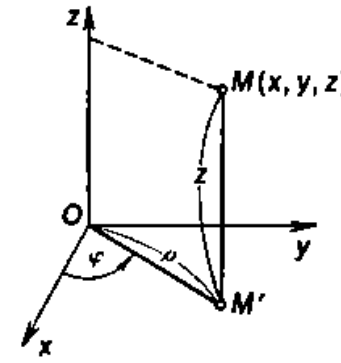



Fig. 197

from rectangular coordinates to new systems of coordinates, the *cylindrical* and *spherical coordinates* being most often used.

A change of variables in a triple integral is carried out according to the following rule.

If a bounded closed domain $V$ of the space $(x, y, z)$ can be mapped one-to-one onto the domain $V^*$ of the space $(u, v, w)$ by means of continuously differentiable functions $x = x(u, v, w)$, $y = y(u, v, w)$, $z = z(u, v, w)$ and the Jacobian $J$ does not vanish in the domain $V^*$, i.e.

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix} \neq 0,$$

then there holds a formula

$$\iiint_V f(x, y, z)\, dx\, dy\, dz$$

$$= \iiint_{V^*} f[x(u, v, w),\ y(u, v, w),\ z(u, v, w)]\, |J|\, du\, dv\, dw.$$

In particular, when we pass from the rectangular coordinates $x$, $y$, $z$ to the cylindrical coordinates $\rho$, $\varphi$, $z$ (Fig. 197) connected

with $x$, $y$, $z$ by the formulas

$$x = \rho\cos\varphi, \quad y = \rho\sin\varphi, \quad z = z$$
$$(0 \leqslant \quad < +\infty, \quad 0 \leqslant \varphi \leqslant 2\pi, \; -\infty < z < +\infty),$$

the Jacobian $J = \rho$ and therefore

$$\iiint\limits_V f(x, y, z)\,dx\,dy\,dz$$
$$= \iiint\limits_{V^*} f(\rho\cos\varphi, \; \rho\sin\varphi, \; z)\,\rho\,d\rho\,d\varphi\,dz. \tag{4}$$

The term "cylindrical coordinates" is due to the fact that the coordinate surface $\rho = \text{const}$ (i.e. a surface whose all points have the

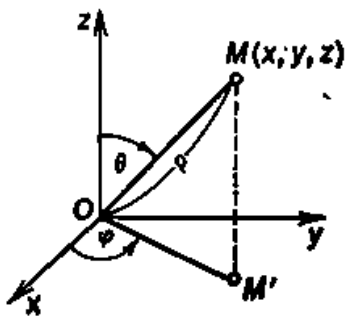



Fig. 198

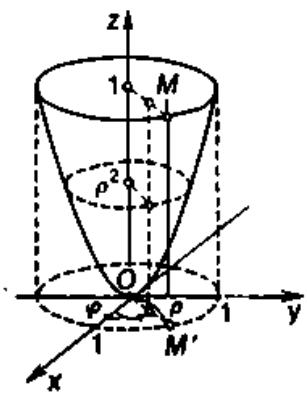



Fig. 199

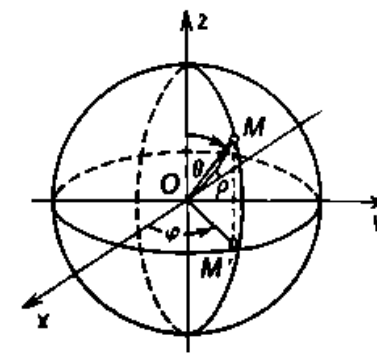



Fig. 200

same coordinate $\rho$) is a cylinder whose rectilinear generatrices are parallel to the $z$-axis.

When we pass from the rectangular coordinates $x$, $y$, $z$ to the spherical coordinates $\rho$, $\varphi$, $\theta$ (Fig. 198) connected with $x$, $y$, $z$ by the formulas

$$x = \rho\sin\theta\cos\varphi, \; y = \rho\sin\theta\sin\varphi, \; z = \rho\cos\theta$$
$$(0 \leqslant \rho < +\infty, \; 0 \leqslant \varphi \leqslant 2\pi, \; 0 \leqslant \theta \leqslant \pi),$$

the Jacobian $J = \rho^2\sin\theta$ and therefore

$$\iiint\limits_V f(x, y, z)\,dx\,dy\,dz$$
$$= \iiint\limits_{V^*} f(\rho\sin\theta\cos\varphi, \; \rho\sin\theta\sin\varphi, \; \rho\cos\theta)\,\rho^2\sin\theta\,d\rho\,d\varphi\,d\theta. \tag{5}$$

The term "spherical coordinates" is due to the fact that the coordinate surface $\rho = \text{const}$ (i.e. a surface whose all points have the same coordinate $\rho$) is a sphere. Spherical coordinates are also known as polar coordinates in space.

When we calculate a triple integral by passing to cylindrical or spherical coordinates, we may not depict the domain $V^*$ and write the limits of integration proceeding directly from the form of the domain $V$, using the geometric meaning of the new coordinates.

**Example 3.** Calculate the integral $\iiint\limits_V (x^2 + y^2)\,dx\,dy\,dz$ by passing to the cylindrical coordinates $x = \rho\cos\varphi$, $y = \rho\sin\varphi$, $z = z$, where $V$ is a domain bounded by the surfaces $z = x^2 + y^2$ and $z = 1$ (Fig. 199).

*Solution.* Since the domain $V$ is projected onto the $xy$-plane into the circle $x^2 + y^2 \leqslant 1$, the coordinate $\varphi$ varies in the limits from 0 to $2\pi$ and the coordinate $\rho$ varies from $\rho = 0$ to $\rho = 1$. The constant values of $\rho$ $(0 \leqslant \rho \leqslant 1)$ in the $xyz$-space is associated with the cylinder $x^2 + y^2 = \rho^2$. Considering the intersection of the cylinder with the domain $V$, we get the variation of the coordinate $z$ from the values for the points lying on the paraboloid $z = x^2 + y^2$ to the values of the points lying on the plane $z = 1$, i.e. from $z = \rho^2$ to $z = 1$. Using formula (4), we get

$$\iiint\limits_V (x^2+y^2)\,dx\,dy\,dz = \int_0^{2\pi} d\varphi \int_0^1 d\rho \int_{\rho^2}^1 \rho^2\cdot\rho\,dz$$

$$= \int_0^{2\pi} d\varphi \int_0^1 [\rho^3 z]_{\rho^2}^1\,d\rho = \int_0^{2\pi} \left[\frac{\rho^4}{4} - \frac{\rho^6}{6}\right]_0^1 d\varphi = \frac{1}{12}\int_0^{2\pi} d\varphi$$

$$= \frac{1}{12}\varphi\Big|_0^{2\pi} = \frac{\pi}{6}.$$

It is difficult to give a definite recommendation as to the use of this or that system of coordinates. It may depend on the domain of integration as well as on the form of the integrand function. However, formula (5), for instance, is more convenient when $f(x, y, z)$ has the form $f(x^2 + y^2 + z^2)$ and also when the domain $V$ is a sphere $x^2 + y^2 + z^2 \leqslant R^2$ or its part.

**Example 4.** Calculate the integral $\iiint\limits_V (x^2 + y^2 + z^2)\,dx\,dy\,dz$, where $V$ is a sphere $x^2 + y^2 + z^2 \leqslant R^2$ (Fig. 200).

*Solution.* In this case it is convenient to pass to spherical coordinates $x = \rho\sin\theta\cos\varphi$, $y = \rho\sin\theta\sin\varphi$, $z = \rho\cos\theta$. It follows from the form of the domain $V$ that the coordinates $\rho$, $\varphi$, $\theta$ vary in the following limits: $\rho$ from 0 to $R$, $\varphi$ from 0 to $2\pi$, $\theta$ from 0 to $\pi$. Since the integrand function

$$x^2 + y^2 + z^2 = \rho^2\sin^2\theta\cos^2\varphi + \rho^2\sin^2\theta\sin^2\varphi + \rho^2\cos^2\theta$$
$$= \rho^2\sin^2\theta + \rho^2\cos^2\theta = \rho^2,$$

we get from formula (5)

$$\iiint\limits_V (x^2+y^2+z^2)\,dx\,dy\,dz=\int\limits_0^R d\rho\int\limits_0^\pi d\theta\int\limits_0^{2\pi}\rho^2\rho^2\sin\theta\,d\varphi$$

$$=\int\limits_0^R\rho^4\,d\rho\int\limits_0^\pi\sin\theta\,d\theta\int\limits_0^{2\pi}d\varphi=2\pi\int\limits_0^R\rho^4\,d\rho\int\limits_0^\pi\sin\theta\,d\theta$$

$$=4\pi\int\limits_0^R\rho^4\,d\rho=4\pi\left.\frac{\rho^5}{5}\right|_0^R=\frac{4\pi R^5}{5}.$$

**4. Some applications of triple integrals.** We shall briefly discuss typical problems of application of triple integrals and give the necessary formulas without deriving them since their derivation is similar to that of the corresponding formulas for double integrals.

If there is a body $V$ with density $\rho(M)=\rho(x, y, z)$ which is a continuous function, then $\iiint\limits_V \rho(x, y, z)\,dx\,dy\,dz$ is the mass $m$ of this body.

The moments of inertia of the body $V$ with density $\rho(M)=\rho(x, y, z)$ about the coordinate axes are defined by the following formulas:

$$I_z=\iiint\limits_V (x^2+y^2)\,\rho(M)\,dv,$$

$$I_y=\iiint\limits_V (x^2+z^2)\,\rho(M)\,dv,$$

$$I_x=\iiint\limits_V (y^2+z^2)\,\rho(M)\,dv.$$

The moment of inertia about the origin

$$I_0=\iiint\limits_V (x^2+y^2+z^2)\,\rho(M)\,dv.$$

The coordinates of the centre of mass are defined by the following formulas:

$$x_c=\frac{\iiint\limits_V x\rho(M)\,dv}{m},\quad y_c=\frac{\iiint\limits_V y\rho(M)\,dv}{m},\quad z_c=\frac{\iiint\limits_V z\rho(M)\,dv}{m},$$

where $x_c$, $y_c$, $z_c$ are the coordinates of the centre of mass and $m$ is the mass of the body.

In particular, if the body in question is homogeneous, i.e. $\rho(x, y, z)=\text{const}$, then the expressions for the coordinates of the

centre of mass become simpler in form

$$x_c = \frac{\iiint\limits_V x\,dv}{v}, \quad y_c = \frac{\iiint\limits_V y\,dv}{v}, \quad z_c = \frac{\iiint\limits_V z\,dv}{v},$$

where $v$ is the volume of the body.

We have pointed out that the triple integral $\iiint\limits_V dx\,dy\,dz$ is equal to the volume of the body $V$. In some cases triple integrals are more convenient for calculating volumes than double integrals since they make it possible to calculate the volume of not only curvilinear cylinders but also of other bodies.

**Example 5.** Find the coordinates of the centre of mass of the upper half of a homogeneous ball $V$ of radius $R$ with centre at the origin.

*Solution.* The half-ball is bounded by the surfaces $z = \sqrt{R^2 - x^2 - y^2}$ and $z = 0$. Since the half-ball is symmetric, $x_c = y_c = 0$. The coordinate $z_c$ can be calculated by the formula

$$z_c = \frac{\iiint\limits_V z\,dx\,dy\,dz}{\iiint\limits_V dx\,dy\,dz} = \frac{\iiint\limits_V z\,dx\,dy\,dz}{\frac{2}{3}\pi R^3}.$$

Passing to the spherical coordinates, we have

$$z_c = \frac{\int\limits_0^{2\pi} d\varphi \int\limits_0^{\pi/2} \sin\theta\cos\theta\,d\theta \int\limits_0^R \rho^3\,d\rho}{\frac{2}{3}\pi R^3} = \frac{2\pi\,\frac{R^4}{4}\cdot\frac{1}{2}}{\frac{2}{3}\pi R^3} = \frac{3}{8}R.$$

### 2.13.11. Surface Integrals

We shall discuss here integrals of functions defined on the surface, the so-called surface integrals.

The theory of surface integrals is similar in many respects to the theory of line integrals. There are surface integrals of the first and of the second kind.

**1. Definition of a surface integral of the first kind.** Let a bounded function $f(M) = f(x, y, z)$ be defined at the points of a smooth or piecewise smooth* surface $S$. We divide the surface $S$ into $n$ arbitrary parts with areas $\Delta S_1, \Delta S_2, \ldots, \Delta S_n$ (Fig. 201). We choose an arbitrary point $M_i(\xi_i, \eta_i, \zeta_i)$ on each subsurface and form a sum

$$\sum_{i=1}^{n} f(M_i)\,\Delta S_i. \tag{1}$$

* The surface is said to be *smooth* if there is a tangent plane at its every point and upon a passage from point to point the position of the plane continuously changes. A surface which consists of a finite number of smooth pieces which are continuously connected is said to be *piecewise smooth*.

Sum (1) is the *integral sum* for the function $f(M)$ over the surface $S$. We designate the largest of the diameters of the subsurfaces as $\lambda$.

**Definition.** *If the integral sum* (1) *has a limit equal to $I$ as $\lambda \to 0$, then this limit is a surface integral of the first kind of the function $f(x, y, z)$ over the surface $S$ and is denoted by one of the following symbols*:

$$I = \iint_S f(M)\,\mathrm{d}S = \iint_S f(x, y, z)\,\mathrm{d}S.$$

In this case the function $f(x, y, z)$ is said to be *integrable over the surface $S$* and $S$ is the *domain of integration*.

This definition is similar in essence to the definition of the double integral. Therefore the properties of double integrals and the condi-

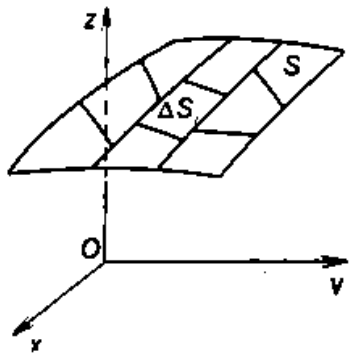



Fig. 201

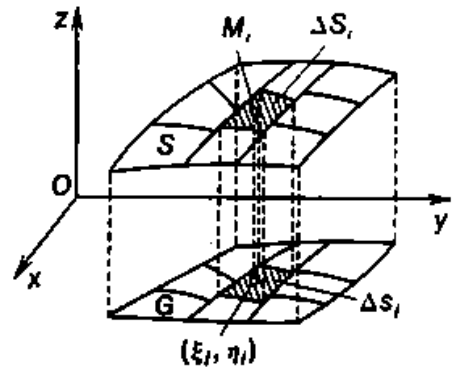



Fig. 202

tions for their existence can be extended to surface integrals without considerable changes.

In particular, if $f(x, y, z) \equiv 1$ on the surface $S$, then

$$\iint_S \mathrm{d}S = \lim_{\lambda \to 0} \sum_{i=1}^{n} \Delta S_i = s,$$

where $s$ is the area of the surface $S$, i.e. using a surface integral of the first kind, we can calculate the area of a surface.

In addition, surface integrals can be used to find masses, static moments, moments of inertia, the coordinates of the centre of mass and other quantities of this kind for material surfaces with a known surface density of distribution of the masses. Problems of this kind can be solved by analogy with the corresponding problems for the case of a material curve, a material plane and a space domain.

**2. Calculating surface integrals of the first kind.** We can calculate a surface integral of the first kind by reducing it to the corresponding double integral.

Let the surface $S$ be defined by the equation $z = z(x, y)$, where the function $z(x, y)$ together with the derivatives $z'_x(x, y)$ and

$z'_y(x, y)$ is continuous in a closed domain $G$ which is the projection of $S$ on the $xy$-plane (Fig. 202), and assume that the function $f(x, y, z)$ is continuous on the surface $S$ and, consequently, is integrable on this surface.

We arbitrarily divide the surface $S$ into $n$ parts and project the division onto the $xy$-plane. We obtain the corresponding division of the domain $G$ into the parts $G_1, G_2, \ldots, G_n$. The area $\Delta S_i$ of each subsurface can be represented as (see formula (2) in 2.13.4, Item 3)

$$\Delta S_i = \iint\limits_{G_i} \sqrt{1 + z_x'^2(x, y) + z_y'^2(x, y)}\, dx\, dy.$$

Applying the mean value theorem to the double integral, we get

$$\Delta S_i = \sqrt{1 + z_x'^2(\xi_i, \eta_i) + z_y'^2(\xi_i, \eta_i)}\ \Delta s_i, \tag{2}$$

where $(\xi_i, \eta_i)$ is a point of the domain $G_i$ and $\Delta s_i$ is the area of $G_i$. We designate the point of a subsurface with coordinates $(\xi_i, \eta_i, \zeta_i)$ as $M_i$, where $\zeta_i = z(\xi_i, \eta_i)$ and $(\xi_i, \eta_i)$ is a point which appears in formula (2). We form an integral sum for the function $f(x, y, z)$ over the surface $S$ choosing the points $M_i$ as intermediate points:

$$\sum_{i=1}^{n} f(\xi_i, \eta_i, \zeta_i)\, \Delta S_i$$

$$= \sum_{i=1}^{n} f[\xi_i, \eta_i, z(\xi_i, \eta_i)] \sqrt{1 + z_x'^2(\xi_i, \eta_i) + z_y'^2(\xi_i, \eta_i)}\, \Delta s_i. \tag{3}$$

The right-hand side of the relation includes the integral sum for the double integral of the function $f[x, y, z(x, y)] \times \sqrt{1 + z_x'^2(x, y) + z_y'^2(x, y)}$ continuous in the domain $G$. Therefore, the limit of the right-hand side of (3) is equal to the double integral

$$\iint\limits_{G} f[x, y, z(x, y)] \sqrt{1 + z_x'^2(x, y) + z_y'^2(x, y)}\, dx\, dy$$

as $\lambda \to 0$. Since the function $f(x, y, z)$ is integrable over the surface $S$, the limit of the left-hand side of (3) is equal to the surface integral

$$\iint\limits_{S} f(x, y, z)\, dS$$

as $\lambda \to 0$. Consequently, passing to the limit in (3) as $\lambda \to 0$, we get the required formula

$$\iint\limits_{S} f(x, y, z)\, dS$$

$$= \iint\limits_{G} f[x, y, z(x, y)] \sqrt{1 + z_x'^2(x, y) + z_y'^2(x, y)}\, dx\, dy, \tag{4}$$

which expresses the surface integral of the first kind in terms of a double integral over the projection of the surface $S$ on the $xy$-plane.

We can obtain, by analogy, formulas which express the surface integrals of the first kind in terms of double integrals over the projections of the surface $S$ on the $yz$-plane and $xz$-plane.

**Example 1.** Calculate the integral $\iint\limits_S \sqrt{1+4x^2+4y^2}\,dS$, where $S$ is a part of the paraboloid of revolution $z = 1 - x^2 - y^2$ cut off by the plane $z = 0$ (Fig. 203).

*Solution.* The surface $S$ defined by the equation $z = 1 - x^2 - y^2$ is projected onto the $xy$-plane into the domain $G$ bounded by the circle $x^2 + y^2 = 1$ (the equation of the circle results from the equation of the paraboloid for $z = 0$). Consequently, the domain $G$ is a circle $x^2 + y^2 \leqslant 1$. In this circle the functions $z = 1 - x^2 - y^2$, $z'_x(x, y) = -2x$, $z'_y(x, y) = -2y$ are continuous. From formula (4) we find that

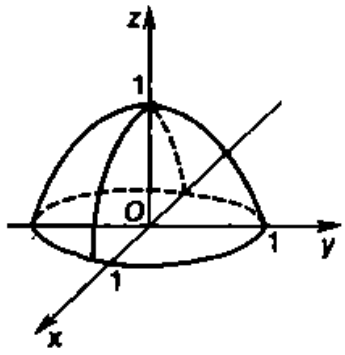



Fig. 203

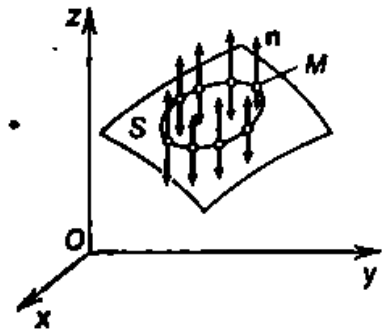



Fig. 204

$$\iint\limits_S f(x, y, z)\,dS = \iint\limits_S \sqrt{1+4x^2+4y^2}\,dS$$

$$= \iint\limits_G \sqrt{1+4x^2+4y^2}\,\sqrt{1+4x^2+4y^2}\,dx\,dy$$

$$= \iint\limits_G (1+4x^2+4y^2)\,dx\,dy.$$

Passing to the polar coordinates $x = \rho\cos\varphi$ and $y = \rho\sin\varphi$ in the double integral obtained, we find that

$$\iint\limits_G (1+4x^2+4y^2)\,dx\,dy = \int\limits_0^{2\pi} d\varphi \int\limits_0^1 (1+4\rho^2)\,\rho\,d\rho$$

$$= \int\limits_0^{2\pi} \left[\frac{\rho^2}{2} + \rho^4\right]_0^1 d\varphi = \frac{3}{2}\int\limits_0^{2\pi} d\varphi = \frac{3}{2}\varphi\Big|_0^{2\pi} = 3\pi.$$

**3. Definition of a surface integral of the second kind.** We first introduce the concept of a side of a surface.

We take an arbitrary point $M$ on a smooth surface $S$ and draw a normal to the surface (vector **n**) through this point. Now we consider, on the surface $S$, a closed contour which passes through the point $M$ and has no points in common with the boundary of the surface $S$. We displace the point $M$ along the closed contour together with the vector **n** so that the vector **n** remains normal to $S$ and its direction varies continuously during the displacement (Fig. 204). The point $M$ returns to its initial position either with the same direction of the normal or with the opposite direction.

If the traverse along any closed contour which lies on the surface $S$ and does not cut its boundary does not change the direction of the normal to the surface on the return to the initial point, then the

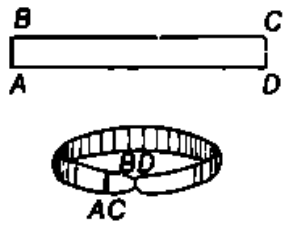



Fig. 205

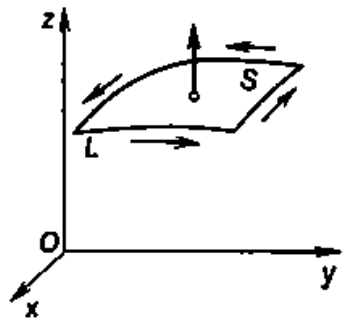



Fig. 206

surface is *two-sided*. A plane, a sphere, any surface defined by the equation $z = f(x, y)$, where $f(x, y)$, $f'_x(x, y)$, $f'_y(x, y)$ are functions continuous in a domain $G$ of the $xy$-plane, can serve as examples of a two-sided surface.

Now if the traverse along any closed contour which lies on the surface $S$ changes the direction of the normal to the opposite on the return to the initial point, then the surface is *one-sided*. The Möbius* strip shown in Fig. 205 is a simple example of a one-sided surface. We can obtain it by taking a strip of paper $ABCD$ and pasting it so that the point $A$ coincides with the point $C$ and the point $B$ with the point $D$, i.e. rotating one of its edges through 180° before pasting the strip. When we traverse the Möbius strip along its centre line and return to the initial point, the direction of the normal changes to the opposite.

In what follows, we shall only consider two-sided surfaces. For a two-sided surface, the collection of all its points with the direction of the normal chosen on them, which changes continuously when we

* Möbius, August Ferdinand (1790-1868), a German mathematician and astronomer.

pass from point to point, is a *side of the surface*, and the choice of a definite side of it is the *orientation of the surface*. A two-sided surface is also known as an *orientable surface* and a one-sided surface as a *nonorientable surface*.

The concept of a side of a surface is closely associated with the concept of the orientation of its boundary.

Let $S$ be an oriented (the side has been chosen) surface bounded by the contour $L$ which has no points of self-intersection. We assume the direction of traverse of the contour $L$ to be *positive* if, during the motion in that direction, the observer, positioned so that the direction of the normal coincides with the vertical direction (from feet to head), sees the surface to his left (Fig. 206). The opposite direction of traverse of the contour is said to be *negative*. If we change the orientation of the surface, i.e. change the direction of the normal to the opposite, then the positive and negative directions of traverse of the contour are interchanged.

Let us define now a surface integral of the second kind.

Assume that $S$ is a smooth surface defined by the equation $z = f(x, y)$, and $R(x, y, z)$ is a bounded function defined at the points of the surface $S$. We choose one of the two sides of the surface, i.e. one of the two possible directions of the normal at the points of the surface (we thus orient the surface). If the normals make acute angles with the $z$-axis, then we say that we have chosen the *upper side* of the surface $z = f(x, y)$, and if they make obtuse angles, then we have chosen the *lower side* of the surface. We arbitrarily divide the surface $S$ into $n$ parts and designate the projection of the $i$th part of the surface on the $xy$-plane as $G_i$. We choose an arbitrary point $M_i(\xi_i, \eta_i, \zeta_i)$ on each subsurface and form a sum

$$\sum_{i=1}^{n} R(\xi_i, \eta_i, \zeta_i)\,\Delta s_i, \tag{5}$$

where $\Delta s_i$ is the area of $G_i$, taken with the plus sign if we have chosen the upper side of the surface $S$, and with the minus sign if we have chosen the lower side of the surface $S$. Sum (5) is the *integral sum* for the function $R(M) = R(x, y, z)$. We designate the largest of the diameters of the subsurfaces as $\lambda$.

**Definition.** *If the integral sum* (5) *has a limit equal to $I$ as $\lambda \to 0$, then this limit is a surface integral of the second kind of the function $R(x, y, z)$ over the chosen side of the surface $S$ and is denoted by one of the following symbols*:

$$I = \iint_S R(M)\,dx\,dy = \iint_S R(x, y, z)\,dx\,dy.$$

In this case the function $R(x, y, z)$ is said to be *integrable over the surface $S$* with respect to the variables $x$ and $y$.

We can similarly define a surface integral of the second kind over the chosen side of the surface $S$ with respect to the variables $y$ and $z$ ($z$ and $x$) of the function $P(x, y, z)$ ($Q(x, y, z)$), which is defined on the surface $S$:

$$\iint_S P(x, y, z)\,dy\,dz \quad \left(\iint_S Q(x, y, z)dz\,dx\right).$$

The sum

$$\iint_S P(x, y, z)\,dy\,dz + \iint_S Q(x, y, z)dz\,dx + \iint_S R(x, y, z)dx\,dy$$

is the *general surface integral of the second kind* and is designated as

$$\iint_S P(x, y, z)\,dy\,dz + Q(x, y, z)dz\,dx + R(x, y, z)\,dx\,dy. \qquad (6)$$

The surface integral of the second kind possesses the same properties as the surface integral of the first kind, but, as distinct from the latter, it changes sign when the side of the surface is changed (reorientation).

The problem of the flux of a vector field, for instance, which will be considered in 2.13.14, leads to the concept of the surface integral of the second kind.

For a one-sided surface the concept of the surface integral of the second kind is not introduced.

**4. Calculating surface integrals of the second kind.** We can calculate a surface integral of the second kind by reducing it to the corresponding double integral.

Assume that the oriented (we choose the upper side) smooth surface $S$ is defined by the equation $z = f(x, y)$, where the function $f(x, y)$ is defined in a closed domain $G$ which is the projection of the surface $S$ on the $xy$-plane, and $R(x, y, z)$ is a continuous function on the surface $S$.

We arbitrarily divide the surface $S$ into $n$ parts and project the division onto the $xy$-plane (Fig. 207). The domain $G$ will respectively be divided into the parts $G_1, G_2, \ldots, G_n$. We choose an arbitrary point $M_i(\xi_i, \eta_i, \zeta_i)$ on each subsurface and form an integral sum $\sum_{i=1}^{n} R(\xi_i, \eta_i, \zeta_i)\,\Delta s_i$, where $\Delta s_i$ is the area of $G_i$. Since $\zeta_i = f(\xi_i, \eta_i)$, it follows that

$$\sum_{i=1}^{n} R(\xi_i, \eta_i, \zeta_i)\Delta s_i = \sum_{i=1}^{n} R[\xi_i, \eta_i, f(\xi_i, \eta_i)]\,\Delta s_i. \qquad (7)$$

The right-hand side of the relation includes the integral sum for the double integral of the function $R[x, y, f(x, y)]$ continuous in

the domain $G$. Passing to the limit in (7) as $\lambda \to 0$, we get the required formula

$$\iint_S R(x, y, z)\,dx\,dy = \iint_G R[x, y, f(x, y)]\,dx\,dy, \tag{8}$$

which expresses the surface integral of the second kind with respect to the variables $x$ and $y$ in terms of a double integral. In addition, formula (8) proves the existence of a surface integral of the function

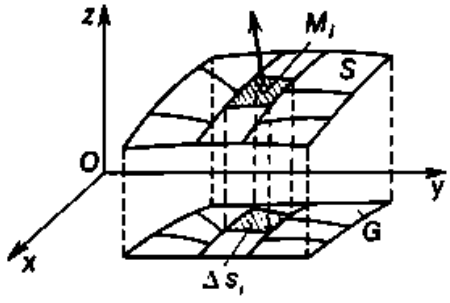



Fig. 207

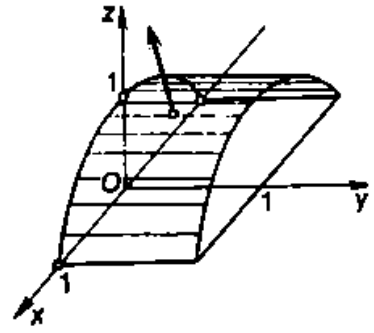



Fig. 208

$R(x, y, z)$ continuous on the surface $S$ being considered. If we choose the lower side of the surface, then the minus sign appears before the integral on the right-hand side of (8).

We can similarly establish the validity of the formulas

$$\iint_S P(x, y, z)\,dy\,dz = \iint_{G_1} P[f(y, z), y, z]\,dy\,dz, \tag{9}$$

$$\iint_S Q(x, y, z)\,dz\,dx = \iint_{G_2} Q[x, f(x, z), z]\,dz\,dx, \tag{10}$$

where the surface $S$ is defined by the equations $x = f(y, z)$ and $y = f(x, z)$, respectively, and $G_1$ and $G_2$ are the projections of the surface $S$ on the $yz$-plane and $xz$-plane respectively.

To calculate the integral of the general type (6), use is made of the same formulas (8)-(10) if the surface $S$ can be projected one-to-one onto the three coordinate axes. In more complicated cases the surface $S$ is divided into parts which possess the indicated properties, and integral (6) is represented as the sum of the integrals over those parts.

**Example 2.** Calculate the integral $\iint_S (y^2 + z^2)\,dx\,dy$, where $S$ is the upper side of the surface $z = \sqrt{1 - x^2}$ cut off by the planes $y = 0$ and $y = 1$ (Fig. 208).

*Solution.* The projection $G$ of the surface $S$ on the $xy$-plane is a rectangle defined by the inequalities $-1 \leqslant x \leqslant 1$ and $0 \leqslant y \leqslant 1$.

From formula (8) we find that

$$\iint_S (y^2+z^2)\,dx\,dy = \iint_G [y^2+(\sqrt{1-x^2})^2]\,dx\,dy$$

$$= \int_{-1}^{1} dx \int_0^1 (y^2+1-x^2)\,dy = \int_{-1}^{1} \left[\frac{y^3}{3}+y-x^2y\right]_0^1 dx$$

$$= \int_{-1}^{1} \left(\frac{4}{3}-x^2\right) dx = \left[\frac{4}{3}x-\frac{x^3}{3}\right]_{-1}^{1} = 2.$$

**Example 3.** Calculate the integral $\iint_S x\,dy\,dz + y\,dz\,dx + z\,dx\,dy$, where $S$ is the upper side of the part of the plane $x+z-1=0$ cut off by the planes $y=0$ and $y=4$ and lying in the first octant (Fig. 209).

*Solution.* By definition,

$$\iint_S x\,dy\,dz + y\,dz\,dx + z\,dx\,dy$$

$$= \iint_{G_1} x(y,z)\,dy\,dz + \iint_S y\,dz\,dx + \iint_{G_2} z(x,y)\,dx\,dy.$$

Here $G_1$ and $G_2$ are the projections of the surface $S$ on the $yz$-plane and $xy$-plane, and $\iint_S y\,dz\,dx = 0$ since the surface $S$ is parallel to the $y$-axis. From formulas (8) and (9) we find that

$$\iint_S z\,dx\,dy = \iint_{G_2} (1-x)\,dx\,dy = \int_0^4 dy \int_0^1 (1-x)\,dx = 2,$$

$$\iint_S x\,dy\,dz = \iint_{G_1} (1-z)\,dy\,dz = \int_0^4 dy \int_0^1 (1-z)\,dz = 2$$

respectively. Consequently,

$$\iint_S x\,dy\,dz + y\,dz\,dx + z\,dx\,dy = 2+0+2 = 4.$$

**5. The relation between surface integrals of the first and the second kind.** We can introduce surface integrals of the second kind in another way, namely, as surface integrals of the first kind in which some special expressions are under the integral sign. We designate the direction cosines of the normal to the oriented surface at its arbitrary point as $\cos\alpha$, $\cos\beta$, $\cos\gamma$. Surface integrals of the second kind differ by their relation to the coordinate planes.

(1) The surface integral of the second kind for the $xy$-plane of the function $R(x, y, z)$ can be expressed in terms of the surface integral of the first kind by means of the formula

$$\iint_S R(x, y, z)\,dx\,dy = \iint_S R(x, y, z)\cos\gamma\,dS. \tag{11}$$

(2) The surface integral of the second kind for the $xz$-plane of the function $Q(x, y, z)$ can be expressed in terms of the surface integral of the first kind by means of the formula

$$\iint_S Q(x, y, z)\,dz\,dx = \iint_S Q(x, y, z)\cos\beta\,dS. \tag{12}$$

(3) The surface integral of the second kind for the $yz$-plane of the

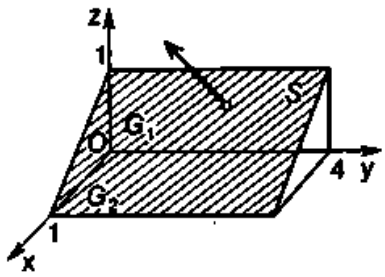



Fig. 209

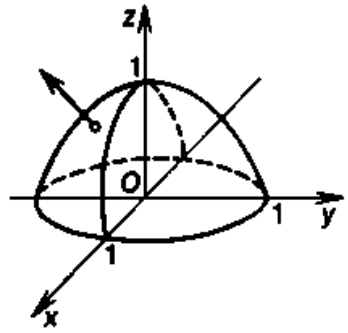



Fig. 210

function $P(x, y, z)$ can be expressed in terms of the surface integral of the first kind by means of the formula

$$\iint_S P(x, y, z)\,dy\,dz = \iint_S P(x, y, z)\cos\alpha\,dS. \tag{13}$$

Summing up formulas (11)-(13), we get a formula which expresses the general surface integral of the second kind over the chosen side of the surface in terms of the surface integral of the first kind

$$\iint_S P\,dy\,dz + Q\,dz\,dx + R\,dx\,dy$$

$$= \iint_S (P\cos\alpha + Q\cos\beta + R\cos\gamma)\,dS. \tag{14}$$

If we choose the other side of the surface, then the direction cosines of the normal $\cos\alpha$, $\cos\beta$, $\cos\gamma$ will change sign and, consequently, the surface integral of the second kind will also change sign.

**Example 4.** Calculate the integral $\iint_S z\cos\gamma\,dS$, where $S$ is the outer side of the hemisphere $x^2 + y^2 + z^2 = 1$ lying above the

$xy$-plane and $\gamma$ is the acute angle between the normal to the surface $S$ and the $z$-axis (Fig. 210).

*Solution.* From formula (11), which connects surface integrals of both kinds, we have

$$\iint_S z\cos\gamma\,dS=\iint_S z\,dx\,dy.$$

The projection $G$ of the surface $S$ on the $xy$-plane is a circle $x^2+y^2\leqslant 1$. From formula (8) we have

$$\iint_S z\,dx\,dy=\iint_G \sqrt{1-x^2-y^2}\,dx\,dy.$$

Passing to polar coordinates in the double integral, we get

$$\iint_G \sqrt{1-x^2-y^2}\,dx\,dy=\int_0^{2\pi} d\varphi\int_0^1 \sqrt{1-\rho^2}\,\rho\,d\rho$$

$$=\int_0^{2\pi}\left[-\frac{(1-\rho^2)^{3/2}}{3}\right]_0^1 d\varphi=\frac{1}{3}\int_0^{2\pi} d\varphi=\frac{2}{3}\pi.$$

### 2.13.12. Ostrogradsky's Theorem

Ostrogradsky's* theorem establishes a relationship between a surface integral taken over a closed surface and a triple integral over a space domain bounded by that surface. This is an analogue of Green's theorem which, as we know, relates a line integral taken over a closed curve and a double integral taken over a plane domain bounded by that curve. Ostrogradsky's theorem is widely used in the analysis itself and in its applications.

We introduce this theorem for a closed domain in space whose boundary cuts any straight line which is parallel to the coordinate axes at no more than two points. For simplicity, we shall call domains of this kind *simple domains*. We shall consider the outer side of the surface which bounds this domain. We assume the surface to be smooth or piecewise smooth.

**Theorem 13.8.** *Let $V$ be a simple closed domain bounded by the surface $S$ and assume that the functions $P(x, y, z)$, $Q(x, y, z)$, $R(x, y, z)$ are continuous together with their partial derivatives of the first order in this domain. Then there holds a formula*

$$\iiint_V \left(\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}+\frac{\partial R}{\partial z}\right)dx\,dy\,dz=\iint_S P\,dy\,dz+Q\,dz\,dx+R\,dx\,dy, \tag{1}$$

*which is known as Ostrogradsky's formula.*

* Ostrogradsky, Mikhail Vasilyevich (1801-1861), a Russian mathematician.

*Proof*. Assume that the domain $G$ is the projection of the surface $S$ (and the domain $V$) on the $xy$-plane (Fig. 211), and $z = z_1(x, y)$ and $z = z_2(x, y)$ are the equations of the corresponding parts of the surface $S$, of the lower part $S_1$ and of the upper part $S_2$.

We reduce the triple integral

$$\iiint\limits_V \frac{\partial R}{\partial z}\,dx\,dy\,dz$$

to a surface integral. For that purpose, we reduce it to an iterated integral and integrate it with respect to $z$ using the Newton-Leibniz formula. We obtain

$$\iiint\limits_V \frac{\partial R}{\partial z}\,dx\,dy\,dz = \iint\limits_G dx\,dy \int\limits_{z_1(x,\,y)}^{z_2(x,\,y)} \frac{\partial R}{\partial z}\,dz$$

$$= \iint\limits_G R[x, y, z_2(x, y)]\,dx\,dy - \iint\limits_G R[x, y, z_1(x, y)]\,dx\,dy.$$

Since the domain $G$ is the projection of both the surface $S_2$ and the surface $S_1$ on the $xy$-plane, we can replace the double integrals by

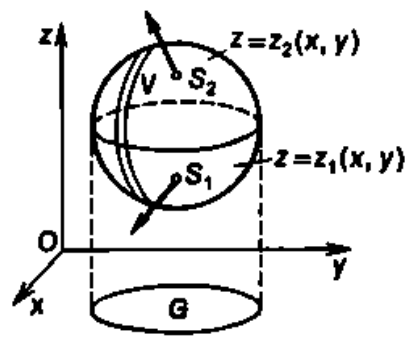



Fig. 211

the surface integrals equal to them (see formula (8) in 2.13.11, Item 4), taken over the upper side of the surface $z = z_2(x, y)$ and over the upper side of the surface $z = z_1(x, y)$ respectively, i.e.

$$\iiint\limits_V \frac{\partial R}{\partial z}\,dx\,dy\,dz = \iint\limits_{S_2} R(x, y, z)\,dx\,dy - \iint\limits_{S_1} R(x, y, z)\,dx\,dy.$$

Changing the side of the surface in the integral taken over $S_1$, we obtain

$$\iiint\limits_V \frac{\partial R}{\partial z}\,dx\,dy\,dz = \iint\limits_{S_2} R(x, y, z)\,dx\,dy$$

$$+ \iint\limits_{S_1} R(x, y, z)\,dx\,dy = \iint\limits_S R\,dx\,dy, \qquad (2)$$

where $S$ is the outer side of the surface which bounds the domain $V$.

We can prove the following formulas by analogy:

$$\iiint_V \frac{\partial P}{\partial x}\,dx\,dy\,dz = \iint_S P\,dy\,dz, \tag{3}$$

$$\iiint_V \frac{\partial Q}{\partial y}\,dx\,dy\,dz = \iint_S Q\,dz\,dx. \tag{4}$$

Summing up relations (2), (3) and (4) term by term, we arrive at formula (1). ■

**Remark.** Ostrogradsky's formula holds true for any closed domain $V$ in space, which can be divided into a finite number of simple domains. Indeed, applying formula (1) to each of the subdomains and adding the results together, we get a triple integral over the whole domain $V$ on the left-hand side of the relation and a surface integral over the surface $S$ bounding the domain $V$ on the right-hand side since the surface integrals over auxiliary surfaces are taken twice over the opposite sides and mutually cancel out when the summation is carried out.

It is convenient to use Ostrogradsky's formula to calculate surface integrals over closed surfaces.

**Example 1.** Calculate the integral $\iint_S x\,dy\,dz + y\,dz\,dx + z\,dx\,dy$, where $S$ is the outer side of a pyramid bounded by the planes $x + y + z = 1$, $x = 0$, $y = 0$, $z = 0$ (see Fig. 196).

*Solution.* Using Ostrogradsky's formula, we obtain

$$\iint_S x\,dy\,dz + y\,dz\,dx + z\,dx\,dy = \iiint_V (1+1+1)\,dx\,dy\,dz$$

$$= 3\iiint_V dx\,dy\,dz = 3\int_0^1 dx \int_0^{1-x} dy \int_0^{1-x-y} dz$$

$$= 3\int_0^1 dx \int_0^{1-x} [z]_0^{1-x-y}\,dy = 3\int_0^1 \left[y - xy - \frac{y^2}{2}\right]_0^{1-x} dx$$

$$= 3\int_0^1 \left[1 - x - x(1-x) - \frac{(1-x)^2}{2}\right] dx = 3\cdot\frac{1}{6} = \frac{1}{2}.$$

**Example 2.** Calculate the integral $\iint_S x^3\,dy\,dz + y^3\,dz\,dx + z^3\,dx\,dy$, where $S$ is the outer side of the sphere $x^2 + y^2 + z^2 = R^2$.

*Solution.* Using Ostrogradsky's formula, we obtain

$$\iint_S x^3\,dy\,dz + y^3\,dz\,dx + z^3\,dx\,dy = 3\iiint_V (x^2 + y^2 + z^2)\,dx\,dy\,dz,$$

whence, introducing spherical coordinates, we obtain

$$3\iiint\limits_V (x^2+y^2+z^2)\,dx\,dy\,dz$$

$$=3\int\limits_0^{2\pi} d\varphi \int\limits_0^{\pi} \sin\theta\,d\theta \int\limits_0^{R} \rho^4\,d\rho=\frac{12}{5}\pi R^5.$$

As has been pointed out (see 2.13.9, Item 1), Green's formula expressed the area of a domain in terms of a line integral taken over its boundary. Just in the same way we can use Ostrogradsky's formula to get an expression for the volume of the domain in the form of a surface integral taken over a closed surface $S$ which is the boundary of that domain. Indeed, we choose functions $P$, $Q$, $R$ such that

$$\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}+\frac{\partial R}{\partial z}=1.$$

Then we get

$$v=\iiint\limits_V dx\,dy\,dz=\iint\limits_S P\,dy\,dz+Q\,dz\,dx+R\,dx\,dy,$$

where $v$ is the volume bounded by the surface $S$.

In particular, setting $P=x/3$, $Q=y/3$, $R=z/3$, we obtain a formula for calculating the volume:

$$v=\frac{1}{3}\iint\limits_S x\,dy\,dz+y\,dz\,dx+z\,dx\,dy.$$

### 2.13.13. Stokes' Theorem

Stokes'* theorem establishes a relationship between surface and line integrals. Just as Green's and Ostrogradsky's formulas, Stokes' formula is widely used in the analysis itself and in its applications.

Assume that $S$ is a surface defined by the equation $z=z(x, y)$, where the function $z(x, y)$, $z'_x(x, y)$, $z'_y(x, y)$ are continuous in the closed domain $G$ which is the projection of $S$ on the $xy$-plane, $L$ is the contour which bounds $S$, and $l$ is its projection on the $xy$-plane, which is the contour bounding the domain $G$. We choose the upper side of the surface $S$ (Fig. 212). Then, under the assumptions made, the following theorem holds true.

**Theorem 13.9.** *If the function $P(x, y, z)$ is continuous together with its partial derivatives of the first order on the surface $S$, then there holds*

* Stokes, Sir George Gabriel (1819-1903), a British physicist and mathematician.

*a formula*

$$\oint_L P(x, y, z)\,dx = \iint_S \left(\frac{\partial P}{\partial z}\cos\beta - \frac{\partial P}{\partial y}\cos\gamma\right) dS, \qquad (1)$$

*where* $\cos\beta$ *and* $\cos\gamma$ *are the direction cosines of the normal to the surface* $S$ *and the contour* $L$ *is traversed in the positive direction.*

*Proof.* We reduce the line integral

$$\oint_L P(x, y, z)\,dx$$

taken along the contour $L$ to the integral taken over the surface $S$. We carry out this transformation according to the scheme

$$\oint_L \to \oint_l \to \iint_G \to \iint_S,$$

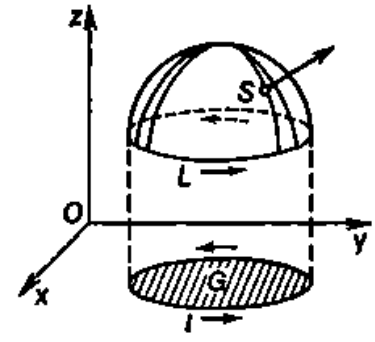


Fig. 212

i.e. we first reduce the line integral along the space contour $L$ to a line integral along the plane contour $l$ and then transform it into a double integral over the domain $G$ and, finally, transform the latter into an integral over the surface $S$.

Since the contour $L$ lies on the surface $S$, the coordinates of its points satisfy the equation $z = z(x, y)$ and therefore the values of the function $P(x, y, z)$ at the points of the contour $L$ are equal to the values of the function $P[x, y, z(x, y)]$ at the corresponding points of the contour $l$ which is the projection of $L$. As to the corresponding projections of the parts of division of the contours $L$ and $l$ on the $x$-axis, they are coincident. Therefore the integral sums for the line integrals of the second kind of the function $P$ along the contours $L$ and $l$ are also coincident and hence the integrals are equal,

$$\oint_L P(x, y, z)\,dx = \oint_l P[x, y, z(x, y)]\,dx.$$

Next we apply Green's formula and pass to the double integral over the domain $G$. We obtain

$$\oint_l P[x, y, z(x, y)]\,dx = -\iint_G \left(\frac{\partial P}{\partial y} + \frac{\partial P}{\partial z} z'_y\right) dx\,dy.$$

Here the integrand function is equal to the partial derivative with respect to $y$ of the composite function which results from $P(x, y, z)$ after the substitution of $z(x, y)$ for $z$.

Since $S$ is the upper side of the surface, i.e. $\cos\gamma > 0$ ($\gamma$ is the acute angle between the normal and the $z$-axis), the normal has projections $-z'_x$, $-z'_y$, 1. And since the direction cosines of the normal

are proportional to the corresponding projections, we have

$$\frac{\cos\beta}{\cos\gamma}=\frac{-z'_y}{1}=-z'_y.$$

Therefore

$$-\iint\limits_G\left(\frac{\partial P}{\partial y}+\frac{\partial P}{\partial z}z'_y\right)\mathrm{d}x\,\mathrm{d}y=-\iint\limits_G\left(\frac{\partial P}{\partial y}-\frac{\partial P}{\partial z}\frac{\cos\beta}{\cos\gamma}\right)\mathrm{d}x\,\mathrm{d}y.$$

Using now formulas (8) and (11) from 2.13.11, we can reduce thie double integral to a surface integral. Then we get

$$-\iint\limits_G\left(\frac{\partial P}{\partial y}-\frac{\partial P}{\partial z}\frac{\cos\beta}{\cos\gamma}\right)\mathrm{d}x\,\mathrm{d}y$$
$$=-\iint\limits_S\left(\frac{\partial P}{\partial y}\cos\gamma-\frac{\partial P}{\partial z}\cos\beta\right)\mathrm{d}S.$$

And thus we have

$$\oint\limits_L P(x,\ y,\ z)\,\mathrm{d}x=\iint\limits_S\left(\frac{\partial P}{\partial z}\cos\beta-\frac{\partial P}{\partial y}\cos\gamma\right)\mathrm{d}S.\ \blacksquare$$

The validity of the following two formulas can be proved by analogy under the requisite conditions:

$$\oint\limits_L Q(x,\ y,\ z)\,\mathrm{d}y=\iint\limits_S\left(\frac{\partial Q}{\partial x}\cos\gamma-\frac{\partial Q}{\partial z}\cos\alpha\right)\mathrm{d}S, \tag{2}$$

$$\oint\limits_L R(x,\ y,\ z)\,\mathrm{d}z=\iint\limits_S\left(\frac{\partial R}{\partial y}\cos\alpha-\frac{\partial R}{\partial x}\cos\beta\right)\mathrm{d}S. \tag{3}$$

Adding together relations (1), (2) and (3) term by term, we get a formula

$$\oint\limits_L P\,\mathrm{d}x+Q\,\mathrm{d}y+R\,\mathrm{d}z$$
$$=\iint\limits_S\left[\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)\cos\gamma+\left(\frac{\partial R}{\partial y}-\frac{\partial Q}{\partial z}\right)\cos\alpha+\left(\frac{\partial P}{\partial z}-\frac{\partial R}{\partial x}\right)\cos\beta\right]\mathrm{d}S,$$

which is known as *Stokes' formula.*

Using the formula which connects surface integrals (14) from 2.13.11, we can rewrite Stokes' formula as

$$\oint\limits_L P\,\mathrm{d}x+Q\,\mathrm{d}y+R\,\mathrm{d}z=\iint\limits_S\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)\mathrm{d}x\,\mathrm{d}y+\left(\frac{\partial R}{\partial y}-\frac{\partial Q}{\partial z}\right)\mathrm{d}y\,\mathrm{d}z$$
$$+\left(\frac{\partial P}{\partial z}-\frac{\partial R}{\partial x}\right)\mathrm{d}z\,\mathrm{d}x. \tag{4}$$

Stokes' formula can be easily memorized since the first term on the right-hand side is the same expression which is under the sign of the double integral in Green's formula, and the second and third expressions result from it by a cyclic permutation of the coordinates $x$, $y$, $z$ and the functions $P$, $Q$, $R$.

In particular, if the surface $S$ is a domain of the $xy$-plane bounded by a contour $L$, then the integrals over $dz\,dx$ and $dy\,dz$ vanish and Stokes' formula turns into Green's formula.

Stokes' formula makes it possible to calculate line integrals along closed contours by means of surface integrals.

**Example.** Using Stokes' formula, calculate the integral $\oint\limits_L x^2y^3\,dx + dy + z\,dz$, where $L$ is a circle defined by the equations $x^2 + y^2 = 1$, $z = 0$, the surface $S$ is the upper side of the hemisphere $x^2 + y^2 + z^2 = 1$ $(z > 0)$ and the contour $L$ is traversed in the positive direction.

*Solution.* Since

$$\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = -3x^2y^2, \quad \frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z} = 0, \quad \frac{\partial P}{\partial z} - \frac{\partial R}{\partial x} = 0,$$

we find from Stokes' formula (4) that

$$\oint\limits_L x^2y^3\,dx + dy + z\,dz = -3\iint\limits_S x^2y^2\,dx\,dy = -\frac{\pi}{8}.$$

It follows from Stokes' formula that if

$$\frac{\partial Q}{\partial x} = \frac{\partial P}{\partial y}, \quad \frac{\partial R}{\partial y} = \frac{\partial Q}{\partial z}, \quad \frac{\partial P}{\partial z} = \frac{\partial R}{\partial x}, \tag{5}$$

then the line integral along any closed space curve $L$ is zero:

$$\oint\limits_L P\,dx + Q\,dy + R\,dz = 0. \tag{6}$$

And this means that here the line integral does not depend on the choice of the path of integration.

As in the case of a plane curve, conditions (5) are the necessary and sufficient conditions for relation (6) to be satisfied.

When conditions (5) and (6) are fulfilled, the element of integration $P\,dx + Q\,dy + R\,dz$ is the total differential of a function $U(x, y, z)$, i.e.

$$dU = P\,dx + Q\,dy + R\,dz,$$

and, consequently,

$$\int\limits_{(x_0,\,y_0,\,z_0)}^{(x_1,\,y_1,\,z_1)} P\,dx + Q\,dy + R\,dz = U(x_1, y_1, z_1) - U(x_0, y_0, z_0).$$

The validity of this relation can be established in the same way as that of the corresponding formula (4) from 2.13.8 for a function of two variables.

### 2.13.14. Scalar and Vector Fields

The concept of a field is at the basis of many notions of modern physics. The study of the field theory is beyond the scope of this course and we will therefore touch on it only briefly.

In the general case, we say that a field of a quantity $u$ is given in space if at every point of space (or of a part of it) the value of this quantity is defined. Thus, for instance, when we study a flow of gas, we have to investigate the behaviour of several fields: the temperature field (the temperature has a definite value at every point), the pressure field, the velocity field etc.

The field of the quantity $u$ is *stationary* (or *steady-state*) if $u$ does not depend on time $t$, otherwise the field is *nonstationary* (or *non-steady-state*). Thus the quantity $u$ is a function of the point $M$ and time $t$.

In problems from physics we most often deal with scalar and vector quantities, and correspondingly we distinguish two kinds of fields, *scalar* and *vector fields*. For simplicity, we shall consider them to be stationary.

**1. The scalar field.** Let $G$ be a domain on the plane or in space. *If at every point $M$ of $G$ a scalar quantity $u$ is defined, then we say that a scalar field is given in the domain $G$.* The concepts of the scalar field and of the function defined in the domain $G$ coincide. The following terminology is usually used: a scalar field is defined by means of the function $u = F(M)$ which is a *scalar function*. If we introduce the system of coordinates $Oxyz$ in space, then every point $M$ will have definite coordinates $x, y, z$ and the scalar quantity $u$ is a function of these coordinates, i.e. $u = F(M) = F(x, y, z)$.

The temperature field of air in a premise may serve as an example of a scalar field if we regard the temperature as a function of a point. At the points which are closer to the source of heat the temperature is higher and farther from the source it is lower. If the temperature turns out to be the same everywhere, then the scalar field is constant.

**2. The vector field.** We introduce the concept of the vector field by analogy with that of the scalar field, i.e. *if at every point $M$ of $G$ a vector $\mathbf{F}(M)$ is defined, then we say that a vector field is specified in the domain $G$.* The function $\mathbf{F}(M)$ by means of which the vector field is defined is a *vector function*.

The force field of any nature may serve as an example of a vector field. Every point of the domain is associated with a definite vector

which has a numerical value and a direction of the force at that point.

**Example 1.** Find the vector field of the velocities $\mathbf{v}(M)$ of the points of a solid body which rotates with a constant angular velocity $\omega$ about an axis.

*Solution.* The velocity $\mathbf{v}$ of the point $M$ is equal to the vector product $\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$, where $\boldsymbol{\omega}$ is the vector of the angular velocity and $\mathbf{r}$ is the radius vector of the point $M$ of the rotating body about some point of the rotation axis. We assume this stationary point of the axis to be the origin and the rotation axis to be the $z$-axis. Then $\boldsymbol{\omega} = \omega\mathbf{k}$, $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ and, consequently,

$$\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & 0 & \omega \\ x & y & z \end{vmatrix} = \begin{vmatrix} 0 & \omega \\ y & z \end{vmatrix} \mathbf{i} - \begin{vmatrix} 0 & \omega \\ x & z \end{vmatrix} \mathbf{j} + \begin{vmatrix} 0 & 0 \\ x & y \end{vmatrix} \mathbf{k} = -\omega y\mathbf{i} + \omega x\mathbf{j}$$

is the required vector field.

**3. The potential field.** We introduce the concept of the potential field for which purpose we consider a scalar field $F(M)$. *If at every point $M$ of $G$ a vector* grad $F$ *is defined, then the field of this vector is a potential field.* In this case the scalar field itself is the *potential* of the vector field, and the vector which defines the potential field is often called a *potential vector*, i.e. the vector $\mathbf{a}(M)$ is potential if there is a scalar function $F(M)$ such that

$$\mathbf{a} = \operatorname{grad} F = \frac{\partial F}{\partial x}\mathbf{i} + \frac{\partial F}{\partial y}\mathbf{j} + \frac{\partial F}{\partial z}\mathbf{k}. \tag{1}$$

A question arises as to the conditions under which the vector field $\mathbf{a}(M)$ is potential. We have actually touched on this problem in 2.13.7. Let $P$, $Q$, $R$ be the projections of the vector $\mathbf{a}$ on the coordinate axes $Ox$, $Oy$, $Oz$ respectively, i.e.

$$\mathbf{a} = \mathbf{a}(M) = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}.$$

By virtue of relation (1), the vector field $\mathbf{a}(M)$ is potential if there is a function $F(M)$ such that

$$\frac{\partial F}{\partial x} = P, \quad \frac{\partial F}{\partial y} = Q, \quad \frac{\partial F}{\partial z} = R. \tag{2}$$

It was shown in Theorem 13.7 that the expression $P\,dx + Q\,dy + R\,dz$ (where $P$, $Q$, $R$ are continuous functions which have first-order continuous partial derivatives) is the total differential of some function $F(x, y, z)$ if and only if $P$, $Q$, $R$ satisfy the conditions

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}, \quad \frac{\partial Q}{\partial z} = \frac{\partial R}{\partial y}, \quad \frac{\partial R}{\partial x} = \frac{\partial P}{\partial z}. \tag{3}$$

But if $P\,dx + Q\,dy + R\,dz = dF$, then relations (2) also hold true, i.e. conditions (3) mean that the given vector field is poten-

tial. In this case the function $F(x, y, z)$ is a *potential function* of the field.

A gravitational field is an example of a potential field. If a mass $m$ is at the origin, then this mass produces a gravitational field. According to Newton's law, a force $\mathbf{F}(M)$ equal to $k\frac{m}{r^2}$ and directed towards the origin acts at every point $M$ of space on a unit mass placed at this point. Here $r = |OM| = \sqrt{x^2 + y^2 + z^2}$ is the distance from the origin $O$ to the point $M$, and $k$ is the proportionality factor.

Let $x, y, z$ be the coordinates of the point $M$. Then the projections $P, Q, R$ of the force $\mathbf{F}(M)$ are defined as follows:

$$P = |\mathbf{F}| \cos\alpha = \frac{km}{r^2}\left(-\frac{x}{r}\right) = -\frac{kmx}{r^3},$$

$$Q = |\mathbf{F}| \cos\beta = \frac{km}{r^2}\left(-\frac{y}{r}\right) = -\frac{kmy}{r^3},$$

$$R = |\mathbf{F}| \cos\gamma = \frac{km}{r^2}\left(-\frac{z}{r}\right) = -\frac{kmz}{r^3},$$

where $\cos\alpha$, $\cos\beta$, $\cos\gamma$ are the direction cosines of the vector $\mathbf{F}(M)$ and, consequently,

$$\mathbf{F}(M) = -\frac{kmx}{r^3}\mathbf{i} - \frac{kmy}{r^3}\mathbf{j} - \frac{kmz}{r^3}\mathbf{k}.$$

It can be verified that the given vector field is potential and its potential function $u(r) = km/r$.

In conclusion, we shall find the work done by the force $\mathbf{F}(M)$ when a unit mass is displaced from the point $B(x_1, y_1, z_1)$ to the point $C(x_2, y_2, z_2)$.

As is known, the work $A$ is expressed by a line integral

$$A = \int_{BC} P\,dx + Q\,dy + R\,dz,$$

where $P, Q, R$ are the projections of the force $\mathbf{F}(M)$ on the coordinate axes. Since this field of force is potential, the element of integration is the total differential and therefore the integral does not depend on the choice of the path of integration and can be obtained from the formula

$$A = \int_B^C P\,dx + Q\,dy + R\,dz = u(C) - u(B),$$

i.e. the work done by the force $\mathbf{F}(M)$ is equal to the difference of the values of the potential function at the points $C$ and $B$. In this case

$$A = u(r_2) - u(r_1) = \frac{km}{r_2} - \frac{km}{r_1} = km\left(\frac{1}{r_2} - \frac{1}{r_1}\right),$$

where $r_1$ and $r_2$ are the distances from the points $B$ and $C$ to the origin.

Note that the domain in which the gravitational field is defined is the whole space except for the origin.

**4. Problem of the vector flux.** Assume that a vector field $\mathbf{v}(M)$ of the velocities of a liquid is given in space, i.e. the space is filled with a moving liquid whose velocity at every point $M(x, y, z)$ is defined by the vector

$$\mathbf{v}(M) = P(x, y, z)\,\mathbf{i} + Q(x, y, z)\,\mathbf{j} + R(x, y, z)\,\mathbf{k},$$

where $P$, $Q$, $R$ are the projections of the velocity on the coordinate axes. Let $P$, $Q$, $R$ be continuous functions of the coordinates. We calculate the quantity $\Pi$ of the liquid which flows through an oriented surface $S$ bounded by the space curve $L$ in unit time, assuming the liquid density to be $\rho = 1$.

Let

$$\mathbf{n} = \cos\alpha\mathbf{i} + \cos\beta\mathbf{j} + \cos\gamma\mathbf{k}$$

be a unit vector of the normal to the surface $S$ and assume that its direction cosines are continuous functions of the coordinates $x, y, z$ of the points of the surface.

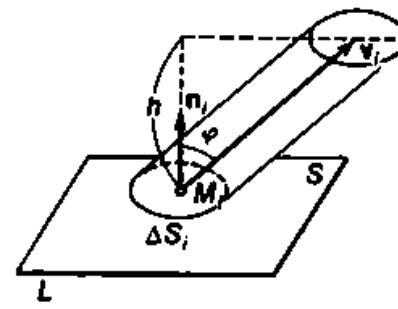


Fig. 213

We divide the surface $S$ into $n$ arbitrary parts with areas $\Delta S_1, \Delta S_2, \ldots, \Delta S_n$ and choose a point $M_i(x_i, y_i, z_i)$ in each of them. We seek the quantity $\Pi_i$ of the liquid flowing through the $i$th part of the surface in unit time (Fig. 213). We designate the angle between the vectors $\mathbf{n}_i = \mathbf{n}(M_i)$ and $\mathbf{v}_i = \mathbf{v}(M_i)$ as $\varphi$. If the angle is acute, i.e. the liquid flows "in the direction" shown by the normal $\mathbf{n}_i$, then we assume the quantity $\Pi_i$ to be positive, and if the angle is obtuse, i.e. the liquid flows in the "reverse direction", then we assume the quantity to be negative.

We can approximately assume that for a sufficiently fine division of the surface $S$ the velocity $\mathbf{v}$ at all points of the $i$th part is constant and equal to $\mathbf{v}(M_i)$ and the subsurfaces are plane. Then the value of $\Pi_i$ is approximately equal to the volume of a cylinder, taken with the corresponding sign, with the base area $\Delta S_i$ and the altitude equal to the absolute value of the projection of the vector $\mathbf{v}_i$ on the normal $\mathbf{n}_i$, i.e. $\Pi_i \cong \Delta S_i h$, where $h$ is the projection. And since $h = |\mathbf{v}_i| \cos\varphi = |\mathbf{v}_i|\,|\mathbf{n}_i| \cos\varphi = (\mathbf{v}_i \cdot \mathbf{n}_i)$, it follows that

$$\Pi_i \cong (\mathbf{v}_i \cdot \mathbf{n}_i)\,\Delta S_i.$$

Summing up with respect to $i$ from 1 to $n$, we obtain an approximate value of the quantity $\Pi$ of the liquid that flows through the oriented

surface $S$ in unit time:

$$\Pi \cong \sum_{i=1}^{n} (\mathbf{v}_i \cdot \mathbf{n}_i)\, \Delta S_i.$$

The sum on the right-hand side is the integral sum for the function $(\mathbf{v} \cdot \mathbf{n})$. Since the projections $P$, $Q$, $R$ of the vector $\mathbf{v}$ and the direction cosines of the vector $\mathbf{n}$ are continuous functions of the coordinates $x$, $y$, $z$ of the points of the surface $S$, the scalar product $\mathbf{v} \cdot \mathbf{n} = P \cos \alpha + Q \cos \beta + R \cos \gamma$ is a continuous function. Consequently, the limit of this sum exists as the largest of the diameters of the subsurfaces tends to zero and is equal to the surface integral of the first kind over the surface $S$ of the function $(\mathbf{v} \cdot \mathbf{n})$. Passing to the limit, we get the exact value of $\Pi$:

$$\Pi = \lim_{\lambda \to 0} \sum_{i=1}^{n} (\mathbf{v}_i \cdot \mathbf{n}_i)\, \Delta S_i = \iint_S (\mathbf{v} \cdot \mathbf{n})\, \mathrm{d}S,$$

or, expressing the scalar product in terms of the coordinates of the vectors,

$$\Pi = \iint_S (P \cos \alpha + Q \cos \beta + R \cos \gamma)\, \mathrm{d}S.$$

Using formula (14) from 2.13.11 which relates surface integrals of the first and the second kind, we finally obtain

$$\Pi = \iint_S P\, \mathrm{d}y\, \mathrm{d}z + Q\, \mathrm{d}z\, \mathrm{d}x + R\, \mathrm{d}x\, \mathrm{d}y. \tag{4}$$

Thus the quantity $\Pi$ of the liquid flowing through the oriented surface $S$ in unit time is a surface integral of the second kind over the chosen side of the surface.

For an arbitrary vector field $\mathbf{v}(M)$ the surface integral of the second kind (4) is known as the *vector flux* $\mathbf{v}(M)$ (or the *flux of the vector field* $\mathbf{v}(M)$, to be more exact) through the surface $S$. For a vector field of some other nature, different from that considered in this example, the flux has naturally some other physical meaning.

**5. Divergence.** Assume that in a domain $V$ bounded by a surface $S$ a vector field $\mathbf{a}(M) = \{P, Q, R\}$ is defined such that the functions $P(M)$, $Q(M)$, $R(M)$ are continuous in $V$ together with their partial derivatives.

**Definition 1.** *The divergence of the vector field* $\mathbf{a}(M)$ *is a scalar function* $\operatorname{div} \mathbf{a}(M)$ *defined by the relation*

$$\operatorname{div} \mathbf{a}(M) = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}. \tag{5}$$

Using the expression for divergence and the concept of a vector flux through a surface, we can write Ostrogradsky's formula (see formula (1) in 2.13.12) in a more concise vector form. The surface integral in Ostrogradsky's formula is a vector flux $\mathbf{a} = \{P, Q, R\}$ through the surface $S$:

$$\iint_S P\,dy\,dz + Q\,dz\,dx + R\,dx\,dy = \iint_S (\mathbf{a}\cdot\mathbf{n})\,dS.$$

Using this expression and relation (5), we write Ostrogradsky's formula in the form

$$\iint_S (\mathbf{a}\cdot\mathbf{n})\,dS = \iiint_V \operatorname{div}\mathbf{a}(M)\,dv. \tag{6}$$

Thus the vector flux $\mathbf{a}(M)$ through the closed surface $S$ is equal to the triple integral of the divergence of the field $\mathbf{a}(M)$ taken over the domain bounded by the surface $S$.

We shall show that the divergence does not depend on the choice of the coordinate system, although its definition was connected with it. For that purpose, we take an arbitrary point $M$, place it into a closed domain $V$ bounded by the surface $S$ and apply Ostrogradsky's formula to the domain $V$. Next, using the mean value theorem for a triple integral, we obtain

$$\iint_S (\mathbf{a}\cdot\mathbf{n})\,dS = \operatorname{div}\mathbf{a}(M_0)\,v,$$

where $M_0$ is a point of the domain $V$ and $v$ is the volume of the domain $V$. Hence

$$\operatorname{div}\mathbf{a}(M_0) = \frac{\iint_S (\mathbf{a}\cdot\mathbf{n})\,dS}{v}.$$

Now we contract the domain $V$ to the point $M$. In this case $v \to 0$, $M_0 \to M$ and we get

$$\operatorname{div}\mathbf{a}(M) = \lim_{v\to 0} \frac{\iint_S (\mathbf{a}\cdot\mathbf{n})\,dS}{v}, \tag{7}$$

i.e. the divergence of the vector field $\mathbf{a}(M)$ at the point $M$ is the limit of the ratio of the vector flux $\mathbf{a}(M)$ through the surface $S$, which envelops the point $M$, to the volume of the domain. And since the flux and the volume are independent of the choice of a coordinate system, the divergence is independent of a coordinate system either, and this is what we wished to show.

We shall use now formula (7) to find the physical meaning of divergence. To do that, we consider the vector field $\mathbf{a}(M)$ to be the

velocity field of a liquid with density $\rho = 1$. We have established in Item 4 that the flux

$$\Pi = \iint_S (\mathbf{a}\cdot\mathbf{n})\,dS$$

of the vector $\mathbf{a}(M)$ is equal to the quantity of the liquid flowing through the surface $S$ in unit time in the direction of the normal $\mathbf{n}$. Let $\mathbf{n}$ be an outer normal. Since $S$ is a closed surface, the flux of the vector $\mathbf{a}(M)$ is evidently equal to the quantity of the liquid which appears or disappears within the domain $V$ bounded by the surface $S$ in unit time. We term that quantity the *total strength of the sources* (if $\Pi > 0$) or the *total strength of the sinks* (if $\Pi < 0$) located in the domain $V$. Let us consider the ratio

$$\frac{\iint_S (\mathbf{a}\cdot\mathbf{n})\,dS}{v}.$$

It expresses the average density of the sources (or sinks), i.e. defines the quantity of the liquid which appears (or disappears) in a unit volume of the domain $V$ in unit time, and the limit

$$\lim_{v\to 0}\frac{\iint_S (\mathbf{a}\cdot\mathbf{n})\,dS}{v},$$

provided that the domain $V$ is contracted to the point $M$, can be termed the *density of the sources* (or *sinks*) at the point $M$. But this limit is equal to $\operatorname{div}\mathbf{a}(M)$. Thus the divergence of the velocity vector field characterizes the density of the sources of the liquid.

If $\operatorname{div}\mathbf{a}(M) > 0$, then, as follows from formula (6), $\Pi > 0$, i.e. there are sources of the liquid in the domain $V$ and the quantity of the outflowing liquid is larger than that of the inflowing one; if $\operatorname{div}\mathbf{a}(M) < 0$, then $\Pi < 0$, i.e. there are sinks in the domain $V$ and the quantity of the inflowing liquid is larger than that of the outflowing one. Now if $\operatorname{div}\mathbf{a}(M) = 0$, then $\Pi = 0$, i.e. there are neither sources nor sinks in the domain $V$ and as much liquid flows into it as flows out of it. This occurs, for instance, in any domain $V$ located in the flux of water in a river.

For an arbitrary vector field $\operatorname{div}\mathbf{a}(M)$ has the same physical meaning: the divergence characterizes the density of the field sources.

The vector field $\mathbf{a}(M)$ is *solenoidal* (a *zero-divergence field*) if $\operatorname{div}\mathbf{a}(M) = 0$ at its every point. We have shown above that an example of such a field is a velocity field of a liquid, when there are no sources or sinks.

**Example 2.** Calculate the divergence of the velocity field $\mathbf{v}(M) = -\omega y\mathbf{i} + \omega x\mathbf{j}$ of a solid body which rotates with constant angular velocity $\omega$ about the $z$-axis.

*Solution.* Here $P = -\omega y$, $Q = \omega x$, $R = 0$ and therefore

$$\operatorname{div} \mathbf{v}(M) = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} = \frac{\partial(-\omega y)}{\partial x} + \frac{\partial(\omega x)}{\partial y} = 0,$$

i.e. this vector field is solenoidal.

**6. Circulation. Rotation.** Assume again that a vector field

$$\mathbf{a}(M) = P(M)\,\mathbf{i} + Q(M)\,\mathbf{j} + R(M)\,\mathbf{k}$$

is defined in some domain and $L$ is a smooth or piecewise smooth curve located in that domain. On the curve $L$ we choose one of the two directions of motion and denote by $\mathbf{dl}$ the vector whose direction at every point is coincident with the direction of motion along the curve at that point and equal, in absolute value, to the differential of the length of the arc:

$$\mathbf{dl} = \mathrm{d}x\mathbf{i} + \mathrm{d}y\mathbf{j} + \mathrm{d}z\mathbf{k}.$$

Then the line integral of the scalar product of the vectors $\mathbf{a}(M)$ and $\mathbf{dl}$

$$\int_L \mathbf{a}(M)\cdot\mathbf{dl} = \int_L P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z$$

is called the *circulation of the vector field* $\mathbf{a}(M)$ *along the curve* $L$. In a force field, circulation expresses the work done by the force field in displacing a particle along the path $L$. Circulation has another physical meaning for fields of some other nature.

**Definition 2.** *The rotation of the vector field* $\mathbf{a}(M)$ *is the vector* $\operatorname{rot}\mathbf{a}(M)$ *defined by the relation*

$$\operatorname{rot}\mathbf{a}(M) = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)\mathbf{i} + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)\mathbf{j} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathbf{k}.$$

Using the notions of the rotation and circulation, we can write Stokes' formula

$$\oint_L P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z$$

$$= \iint_S \left[\left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)\cos\alpha + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)\cos\beta + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\cos\gamma\right]\mathrm{d}S$$

in a more concise vector form

$$\oint_L \mathbf{a}(M)\cdot\mathbf{dl} = \iint_S [\operatorname{rot}\mathbf{a}(M)\cdot\mathbf{n}]\,\mathrm{d}S.$$

Thus the circulation of the vector field $\mathbf{a}(M)$ around a closed contour $L$ is equal to the flux of the rotation of this vector field through the surface $S$ bounded by the contour $L$.

As for divergence, we can show that rot **a** $(M)$ does not depend on the choice of the coordinate system and is defined only by the vector field **a** $(M)$ itself.

**Example 3.** Calculate the rotation of the velocity field $\mathbf{v}(M) = -\omega y\mathbf{i} + \omega x\mathbf{j}$ of a solid body rotating with constant angular velocity $\omega$ about the $z$-axis.

*Solution.* Using the definition of the rotation, we obtain

$$\operatorname{rot}\mathbf{a}(M) = \left(\frac{\partial 0}{\partial y} - \frac{\partial(\omega x)}{\partial z}\right)\mathbf{i} + \left(\frac{\partial(-\omega y)}{\partial z} - \frac{\partial 0}{\partial x}\right)\mathbf{j} + \left(\frac{\partial(\omega x)}{\partial x} - \frac{\partial(-\omega y)}{\partial y}\right)\mathbf{k} = 2\omega\mathbf{k},$$

i.e. the rotation of this vector field is directed along the $z$-axis and is equal, in absolute value, to twice the angular velocity.

The concept of rotation is related to the concept of a potential field. We have shown that the vector field $\mathbf{a}(M) = \{P, Q, R\}$ is potential if and only if

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}, \quad \frac{\partial Q}{\partial z} = \frac{\partial R}{\partial y}, \quad \frac{\partial R}{\partial x} = \frac{\partial P}{\partial z}.$$

But this means that all the three coordinates of the rotation of the field **a** $(M)$ are zero, i.e. for the vector field **a** $(M)$ to be potential, it is necessary and sufficient that the condition

$$\operatorname{rot}\mathbf{a}(M) = 0$$

be satisfied.

**7. The Hamiltonian operator.** It is conventional to represent the main concepts of the field theory, viz. gradient, divergence, rotation, and operations involving them, using the *Hamiltonian**, or "*nabla*" *operator*:

$$\nabla = \frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}.$$

We regard the operator $\nabla$ as a symbolic vector with coordinates $\frac{\partial}{\partial x}$, $\frac{\partial}{\partial y}$, $\frac{\partial}{\partial z}$ and carry out the operations involving it according to the rules of vector algebra. In this case we understand the product of $\frac{\partial}{\partial x}$, $\frac{\partial}{\partial y}$, $\frac{\partial}{\partial z}$ by a scalar function to be a partial derivative of that function with respect to $x$, $y$, $z$ respectively.

**Examples.**

1. Let $u(x, y, z)$ be a scalar function. Then the multiplication of the operator $\nabla$ by the function $u$ yields the gradient of this function, i.e.

$$\nabla u = \left(\frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}\right)u = \frac{\partial u}{\partial x}\mathbf{i} + \frac{\partial u}{\partial y}\mathbf{j} + \frac{\partial u}{\partial z}\mathbf{k} = \operatorname{grad} u.$$

* Hamilton, Sir William Rowan (1805-1865), an Irish mathematician.

2. Let $\mathbf{a}(M) = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ be a vector function. Then the scalar multiplication of the operator $\nabla$ by the vector function $\mathbf{a}(M)$ yields the divergence of this function, i.e.

$$\nabla\cdot\mathbf{a}(M) = \left(\frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}\right)\cdot(P\mathbf{i} + Q\mathbf{j} + R\mathbf{k})$$
$$= \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} = \operatorname{div}\mathbf{a}(M).$$

3. The vector multiplication of the operator $\nabla$ by the vector function $\mathbf{a}(M)$ yields the rotation of this function, i.e.

$$\nabla\times\mathbf{a}(M) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{vmatrix}$$
$$= \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)\mathbf{i} + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)\mathbf{j} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathbf{k} = \operatorname{rot}\mathbf{a}(M).$$

In applications we often come across the so-called second-order operations, i.e. pairwise combinations of the three indicated operations. We shall consider the most important of them.

1°. $\operatorname{div}\operatorname{rot}\mathbf{a}(M) = 0$. Indeed,

$$\operatorname{div}\operatorname{rot}\mathbf{a}(M) = \frac{\partial}{\partial x}\left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) + \frac{\partial}{\partial y}\left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) + \frac{\partial}{\partial z}\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)$$
$$= \frac{\partial^2 R}{\partial x\,\partial y} - \frac{\partial^2 Q}{\partial x\,\partial z} + \frac{\partial^2 P}{\partial y\,\partial z} - \frac{\partial^2 R}{\partial y\,\partial x} + \frac{\partial^2 Q}{\partial z\,\partial x} - \frac{\partial^2 P}{\partial z\,\partial y} = 0$$

by virtue of the equality of the mixed second-order derivatives. We can easily obtain the same result using the operator $\nabla$:

$$\operatorname{div}\operatorname{rot}\mathbf{a}(M) = \nabla\cdot(\nabla\times\mathbf{a}) = 0$$

since we have here a mixed product of three "vectors", $\nabla$, $\nabla$, $\mathbf{a}$, two of which are the same. Such a product is evidently zero.

2°. $\operatorname{rot}\operatorname{grad} u = 0$. Indeed,

$$\operatorname{rot}\operatorname{grad} u = \left[\frac{\partial}{\partial y}\left(\frac{\partial u}{\partial z}\right) - \frac{\partial}{\partial z}\left(\frac{\partial u}{\partial y}\right)\right]\mathbf{i} + \left[\frac{\partial}{\partial z}\left(\frac{\partial u}{\partial x}\right) - \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial z}\right)\right]\mathbf{j}$$
$$+ \left[\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial y}\right) - \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial x}\right)\right]\mathbf{k} = \left(\frac{\partial^2 u}{\partial y\,\partial z} - \frac{\partial^2 u}{\partial z\,\partial y}\right)\mathbf{i}$$
$$+ \left(\frac{\partial^2 u}{\partial z\,\partial x} - \frac{\partial^2 u}{\partial x\,\partial z}\right)\mathbf{j} + \left(\frac{\partial^2 u}{\partial x\,\partial y} - \frac{\partial^2 u}{\partial y\,\partial x}\right)\mathbf{k} = 0$$

by virtue of the equality of the mixed second-order derivatives. We can easily obtain the same result using the operator $\nabla$:

$$\operatorname{rot}\operatorname{grad} u = \nabla\times(\nabla u) = (\nabla\times\nabla)\,u = 0$$

since the vector product of the same "vectors" is zero.

3°. div grad $u=\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}$. Indeed,

$$\operatorname{grad} u=\frac{\partial u}{\partial x}\mathbf{i}+\frac{\partial u}{\partial y}\mathbf{j}+\frac{\partial u}{\partial z}\mathbf{k},$$

$$\operatorname{div}\operatorname{grad} u=\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\right)+\frac{\partial}{\partial y}\left(\frac{\partial u}{\partial y}\right)+\frac{\partial}{\partial z}\left(\frac{\partial u}{\partial z}\right)$$

$$=\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}. \qquad (8)$$

The right-hand side of relation (8) is symbolically designated as

$$\Delta u=\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2} \quad \text{or} \quad \Delta u=\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)u.$$

The symbol

$$\Delta=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}$$

is known as the *Laplacian* operator*, or simply *Laplacian*. It is natural to regard the Laplacian operator $\Delta$ as the scalar square of the "vector" $\nabla$. Indeed,

$$(\nabla\cdot\nabla)=\left(\frac{\partial}{\partial x}\right)^2+\left(\frac{\partial}{\partial y}\right)^2+\left(\frac{\partial}{\partial z}\right)^2=\Delta.$$

Therefore, using the operator $\nabla$, we can write relation (8) as

$$\operatorname{div}\operatorname{grad} u=\nabla\cdot(\nabla u)=(\nabla\cdot\nabla)\,u=\nabla^2 u.$$

Note that the equation

$$\Delta u=0$$

is known as *Laplace's equation*. It is used to describe various stationary processes of different physical natures, say, a stationary heat distribution, an electrostatic field of point charges, the steady-state motion of an incompressible liquid inside a domain etc. The scalar field $u(x, y, z)$ which satisfies the condition $\Delta u=0$ is known as *Laplace's*, or *harmonic*, *field*.

---

* Laplace, Pierre Simon (1749-1827), a French mathematician, astronomer and physicist.

# Part Three

# SERIES. DIFFERENTIAL EQUATIONS

## Chapter 14

## SERIES

In this chapter we shall discuss series which constitute an important mathematical apparatus used for calculations and research in various divisions of mathematics itself and in many of its applications.

### 3.14.1. Number Series

**1. Main definitions.** Consider a number sequence $a_1, a_2, a_3, \ldots, a_n, \ldots$. *An expression of the form*

$$a_1 + a_2 + a_3 + \ldots + a_n + \ldots = \sum_{n=1}^{\infty} a_n \tag{1}$$

*is called a number series or simply a series.*

The numbers $a_1, a_2, \ldots, a_n, \ldots$ are the *members*, or *terms, of the series*, the term $a_n$ with an arbitrary number is the *general term of the series*.

The sums of a finite number of terms of a series

$$S_1 = a_1,\ S_2 = a_1 + a_2,\ S_3 = a_1 + a_2 + a_3, \ldots,$$
$$S_n = a_1 + a_2 + a_3 + \ldots + a_n, \ldots$$

are *partial sums of series* (1). Since the number of terms of a series is infinite, the partial sums of the series form an infinite sequence of partial sums

$$S_1,\ S_2,\ S_3, \ldots, S_n, \ldots. \tag{2}$$

Series (1) is said to be *convergent* if the sequence of its partial sums (2) converges to some number $S$ which in this case is the *sum of series* (1). Symbolically it is written as

$$S = a_1 + a_2 + a_3 + \ldots + a_n + \ldots, \quad \text{or} \quad S = \sum_{n=1}^{\infty} a_n.$$

If the sequence of partial sums (2) diverges, then series (1) is *divergent*.

**Example 1.** Let us show that the series

$$\frac{1}{1\cdot 2}+\frac{1}{2\cdot 3}+\frac{1}{3\cdot 4}+\ldots+\frac{1}{n(n+1)}+\ldots=\sum_{n=1}^{\infty}\frac{1}{n(n+1)}$$

converges. We take the sum $S_n$ of the first $n$ terms of the series

$$S_n=\frac{1}{1\cdot 2}+\frac{1}{2\cdot 3}+\ldots+\frac{1}{n(n+1)}.$$

The terms of this sum can be represented as

$$\frac{1}{1\cdot 2}=1-\frac{1}{2},\quad \frac{1}{2\cdot 3}=\frac{1}{2}-\frac{1}{3},\quad \frac{1}{3\cdot 4}=\frac{1}{3}-\frac{1}{4},\quad \ldots,$$
$$\frac{1}{n(n+1)}=\frac{1}{n}-\frac{1}{n+1}$$

and therefore

$$S_n=\left(1-\frac{1}{2}\right)+\left(\frac{1}{2}-\frac{1}{3}\right)+\left(\frac{1}{3}-\frac{1}{4}\right)+\ldots$$
$$+\left(\frac{1}{n}-\frac{1}{n+1}\right)=1-\frac{1}{n+1}.$$

It follows that the limit of the sequence of the partial sums of the given series is unity:

$$\lim_{n\to\infty}S_n=\lim_{n\to\infty}\left(1-\frac{1}{n+1}\right)=1-\lim_{n\to\infty}\frac{1}{n+1}=1.$$

Thus the series converges and its sum $S$ is unity.

**Example 2.** Let us find whether the following series is convergent or divergent:

$$1-1+1-1+\ldots+(-1)^{n-1}+\ldots=\sum_{n=1}^{\infty}(-1)^{n-1}.$$

The sequence of its partial sums has the form $S_1=1$, $S_2=0$, $S_3=1$, $S_4=0$, ... and, hence, does not converge to any limit, and therefore the series is divergent.

**Example 3.** Let us consider a series composed of the elements of the geometric progression

$$a+aq+aq^2+\ldots+aq^{n-1}+\ldots=\sum_{n=1}^{\infty}aq^{n-1},\quad a\neq 0. \tag{3}$$

For $q\neq 1$ the partial sum $S_n$ of this series has the form

$$S_n=a+aq+aq^2+\ldots+aq^{n-1}=\frac{a-aq^n}{1-q}=\frac{a}{1-q}-\frac{aq^n}{1-q}.$$

Hence

(1) if $|q|<1$, then $\lim\limits_{n\to\infty} S_n=\lim\limits_{n\to\infty}\frac{a}{1-q}-\lim\limits_{n\to\infty}\frac{aq^n}{1-q}=\frac{a}{1-q}$, i.e. the series converges and its sum $S=\frac{a}{1-q}$. For instance, for $a=1$ and $q=1/2$ we have

$$S=1+\frac{1}{2}+\frac{1}{2^2}+\ldots+\frac{1}{2^{n-1}}+\ldots=2;$$

(2) if $|q|>1$, then $\lim\limits_{n\to\infty} S_n=\lim\limits_{n\to\infty}\frac{a-aq^n}{1-q}=\infty$, i.e. the series diverges;

(3) for $q=1$ series (3) assumes the form $a+a+a+\ldots+a+\ldots$. In this case $\lim\limits_{n\to\infty} S_n=\lim\limits_{n\to\infty}\underbrace{(a+a+a+\ldots+a)}_{n\text{ times}}=\lim\limits_{n\to\infty} n\cdot a=\infty$, i.e. the series diverges;

(4) for $q=-1$ series (3) assumes the form $a-a+a-a+\ldots$. For this series $S_n=\frac{a}{2}-\frac{a(-1)^n}{2}$, i.e. $S_n=0$ for an even $n$ and $S_n=a$ for an odd $n$. Consequently, $\lim\limits_{n\to\infty} S_n$ does not exist and the series diverges.

Thus series (3) is convergent for $|q|<1$ and divergent for $|q|\geqslant 1$.

**2. Properties of convergent series.**

**Theorem 14.1.** *If the series*

$$a_1+a_2+a_3+\ldots+a_{k-1}+a_k+a_{k+1}+\ldots+a_{n-1}+a_n+\ldots=\sum_{n=1}^{\infty} a_n \quad (4)$$

*converges, then the series*

$$a_{k+1}+\ldots+a_{n-1}+a_n+\ldots=\sum_{n=k+1}^{\infty} a_n \quad (5)$$

*converges too, and, conversely, if series* (5) *converges, then series* (4) *converges too.*

To put it otherwise, the convergence of a series is not affected by the removal of a finite number of its first terms.

*Proof.* Assume that series (4) converges and has a sum $S$, i.e. $\lim\limits_{n\to\infty} S_n=S$. We denote the sum of the deleted terms of series (4) by $S_k$ and the sum of the first $n-k$ terms of series (5) by $\sigma_{n-k}$. Then we have

$$S_n=S_k+\sigma_{n-k}, \quad (6)$$

where $S_k$ is a number independent of $n$. It follows from relation (6) that

$$\lim_{n\to\infty}\sigma_{n-k}=\lim_{n\to\infty}(S_n-S_k)=\lim_{n\to\infty}S_n-\lim_{n\to\infty}S_k=S-S_k,$$

i.e. the sequence $\{\sigma_{n-k}\}$ of partial sums of series (5) has a limit, and this signifies that series (5) is convergent.

Assume now that series (5) converges and has a sum $\sigma$, i.e. $\lim\limits_{n\to\infty} \sigma_{n-k} = \sigma$. Then it follows from (6) that

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty} (S_k + \sigma_{n-k}) = \lim_{n\to\infty} S_k + \lim_{n\to\infty} \sigma_{n-k} = S_k + \sigma,$$

and this signifies that series (4) is convergent. ■

Ordinary arithmetic operations may involve convergent series.

**Theorem 14.2.** *If the series* $\sum\limits_{n=1}^{\infty} a_n$ *converges and its sum is* $S$, *then the series* $\sum\limits_{n=1}^{\infty} ca_n$, *where* $c$ *is a number, also converges and its sum is* $cS$.

*Proof.* Assume that $S_n$ is a partial sum of the series $\sum\limits_{n=1}^{\infty} a_n$, and $\sigma_n$ is a partial sum of the series $\sum\limits_{n=1}^{\infty} ca_n$. Then

$$\sigma_n = ca_1 + ca_2 + ca_3 + \ldots + ca_n = c(a_1 + a_2 + a_3 + \ldots + a_n) = cS_n.$$

Passing to the limit as $n \to \infty$, we find from this that

$$\lim_{n\to\infty} \sigma_n = \lim_{n\to\infty} cS_n = c \lim_{n\to\infty} S_n = cS,$$

i.e. the sequence $\{\sigma_n\}$ of partial sums of the series $\sum\limits_{n=1}^{\infty} ca_n$ converges to $cS$. Consequently, $\sum\limits_{n=1}^{\infty} ca_n = cS$. ■

**Theorem 14.3.** *If the series* $\sum\limits_{n=1}^{\infty} a_n$ *and* $\sum\limits_{n=1}^{\infty} b_n$ *are convergent and their sums are* $S$ *and* $\sigma$ *respectively, then the series* $\sum\limits_{n=1}^{\infty} (a_n \pm b_n)$ *is also convergent and its sum is* $S \pm \sigma$.

*Proof.* Assume that $S_n$ and $\sigma_n$ are partial sums of the series $\sum\limits_{n=1}^{\infty} a_n$ and $\sum\limits_{n=1}^{\infty} b_n$, and $\tau_n$ is a partial sum of the series $\sum\limits_{n=1}^{\infty} (a_n \pm b_n)$. Then

$$\tau_n = (a_1 \pm b_1) + (a_2 \pm b_2) + \ldots + (a_n \pm b_n)$$
$$= (a_1 + a_2 + \ldots + a_n) \pm (b_1 + b_2 + \ldots + b_n) = S_n \pm \sigma_n.$$

Passing to the limit as $n \to \infty$, we find from this that

$$\lim_{n\to\infty} \tau_n = \lim_{n\to\infty} (S_n \pm \sigma_n) = \lim_{n\to\infty} S_n \pm \lim_{n\to\infty} \sigma_n = S \pm \sigma,$$

i.e. the sequence $\{\tau_n\}$ of partial sums of the series $\sum_{n=1}^{\infty}(a_n \pm b_n)$ converges to $S \pm \sigma$. Consequently, $\sum_{n=1}^{\infty}(a_n \pm b_n) = S \pm \sigma$. ■

We have thus established that convergent series can be multiplied by a number, added term by term and subracted in the same way as finite sums.

**3. The necessary condition for convergence of a series.** Two problems arise when we consider series: (1) we must test the series for convergence and (2) knowing that the series is convergent, find its sum. We shall mainly concern ourselves with the first problem which is theoretic in nature. Here is the necessary condition for convergence of a series.

**Theorem 14.4.** *If the series* $\sum_{n=1}^{\infty} a_n$ *is convergent, then its general term tends to zero, i.e.* $\lim_{n\to\infty} a_n = 0$.

*Proof.* The series $\sum_{n=1}^{\infty} a_n$ is convergent by the hypothesis. We denote its sum by $S$. Then we consider the partial sums of the series $S_n = a_1 + a_2 + \ldots + a_{n-1} + a_n$ and $S_{n-1} = a_1 + a_2 + \ldots + a_{n-1}$. From this we have $a_n = S_n - S_{n-1}$. Since $S_n \to S$ and $S_{n-1} \to S$ as $n \to \infty$, it follows that

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty}(S_n - S_{n-1}) = \lim_{n\to\infty} S_n - \lim_{n\to\infty} S_{n-1} = S - S = 0. \ ■$$

The condition $\lim_{n\to\infty} a_n = 0$ is a necessary but not sufficient condition for convergence of a series.

**Example.** Consider a series

$$1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} + \ldots = \sum_{n=1}^{\infty} \frac{1}{n},$$

which is a *harmonic series*. The necessary condition for convergence is evidently fulfilled for the harmonic series since $\lim_{n\to\infty} a_n = \lim_{n\to\infty} \frac{1}{n} = 0$. We shall prove that this series is divergent. Indeed, were this series convergent, then, denoting its sum by $S$, we would have

$$\lim_{n\to\infty}(S_{2n} - S_n) = \lim_{n\to\infty} S_{2n} - \lim_{n\to\infty} S_n = S - S = 0.$$

But

$$S_{2n}-S_n=\frac{1}{n+1}+\frac{1}{n+2}+\ldots$$

$$+\frac{1}{2n}>\frac{1}{2n}+\frac{1}{2n}+\ldots+\frac{1}{2n}=n\frac{1}{2n}=\frac{1}{2},$$

i.e. $S_{2n}-S_n>1/2$. It follows that the equality $\lim\limits_{n\to\infty}(S_{2n}-S_n)=0$ is impossible, i.e. the harmonic series is divergent.

Thus, if the general term of a series tends to zero, the series is not necessarily convergent. An additional investigation is needed which can be carried out with the aid of the sufficient conditions (tests) for convergence of the series.

But if the general term of some series does not tend to zero, then Theorem 14.4 makes it possible to say at once that the series is divergent.

### 3.14.2. Series with Nonnegative Terms

We shall discuss now some sufficient conditions for convergence of series with nonnegative terms. We shall first prove a theorem which will be used in what follows.

**Theorem 14.5.** *For the series* $\sum\limits_{n=1}^{\infty} a_n$ *with nonnegative terms to be convergent, it is, necessary and sufficient that the sequence of partial sums of the series be bounded.*

*Proof.* **Necessity.** Let the series $\sum\limits_{n=1}^{\infty} a_n$ be convergent. This means that the sequence of its partial sums has a limit. By virtue of Theorem 2.6, every convergent sequence is bounded.

**Sufficiency.** Let the sequence of partial sums of the series $\sum\limits_{n=1}^{\infty} a_n$ be bounded. Since the series $\sum\limits_{n=1}^{\infty} a_n$ has nonnegative terms, its partial sums form a nondecreasing sequence $0\leqslant S_1\leqslant S_2\leqslant\ldots\leqslant S_n\leqslant S_{n+1}\leqslant\ldots$. By virtue of Theorem 2.12 on monotone bounded sequences, it is convergent, i.e. the series $\sum\limits_{n=1}^{\infty} a_n$ is convergent. ■

**The sufficient conditions for convergence of a series.** We shall establish a number of tests which allow us to find whether a series in question is convergent or divergent.

**Comparison test.**

**Theorem 14.6.** *Assume that we are given two series* $\sum_{n=1}^{\infty} a_n$ *and* $\sum_{n=1}^{\infty} b_n$ *with nonnegative terms and the inequality* $a_n \leqslant b_n$ *holds true for all n. Then the convergence of the series* $\sum_{n=1}^{\infty} b_n$ *implies the convergence of the series* $\sum_{n=1}^{\infty} a_n$ *and the divergence of the series* $\sum_{n=1}^{\infty} a_n$ *implies the divergence of the series* $\sum_{n=1}^{\infty} b_n$.

*Proof.* We designate the partial sums of the series $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ as $S_n$ and $\sigma_n$ respectively. It follows from the inequality $a_n \leqslant b_n$ that

$$S_n \leqslant \sigma_n. \tag{7}$$

If the series $\sum_{n=1}^{\infty} b_n$ is convergent, then, according to Theorem 14.5 (necessity), the sequence of its partial sums is bounded, i.e. $\sigma_n \leqslant M$ for any $n$, where $M$ is some number. But then, according to formula (7), $S_n \leqslant M$ as well, whence it follows, according to the same Theorem 14.5 (sufficiency), that the series $\sum_{n=1}^{\infty} a_n$ is convergent.

Now if the series $\sum_{n=1}^{\infty} a_n$ is divergent, then the series $\sum_{n=1}^{\infty} b_n$ is divergent too since, assuming the series $\sum_{n=1}^{\infty} b_n$ to be convergent, we find that the series $\sum_{n=1}^{\infty} a_n$ is convergent in accordance with what we have just proved, and this contradicts the hypothesis of the theorem. ■

**Example 1.** The series $\sum_{n=1}^{\infty} \frac{1}{(n+1)^{n+1}}$ is convergent since the series of the terms of the geometric progression $\sum_{n=1}^{\infty} \frac{1}{2^{n+1}}$ $\left(q=\frac{1}{2}<1\right)$ is convergent and the terms of the given series are not greater than

the respective terms of the series of the convergent geometric progression: $\frac{1}{(n+1)^{n+1}} \leqslant \frac{1}{2^{n+1}}$.

**Example 2.** The series $\sum_{n=1}^{\infty} \frac{1}{n^s}$ $(s \leqslant 1)$ is divergent since its terms are not smaller than the terms of the harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$, i.e. $\frac{1}{n} \leqslant \frac{1}{n^s}$, and the harmonic series is divergent.

There are tests for the convergence of series which make it possible to decide whether the series is convergent or divergent without comparing it to another series about which we know whether it is convergent or divergent. Let us consider two of them.

**D'Alembert's* test.**

**Theorem 14.7.** *Assume that the series* $\sum_{n=1}^{\infty} a_n$ *has positive terms and there is a limit* $\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = \rho$. *Then* (a) *the series is convergent for* $\rho < 1$; (b) *the series is divergent for* $\rho > 1$.

*Proof.* (a) Let $\rho < 1$ and $\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = \rho$. We shall prove that the series $\sum_{n=1}^{\infty} a_n$ is convergent. By the definition of the limit of a number sequence, for any $\varepsilon > 0$ there is a number $N$ such that the inequality $\left|\frac{a_{n+1}}{a_n} - \rho\right| < \varepsilon$ is satisfied for $n \geqslant N$. From this it follows that

$$\rho - \varepsilon < \frac{a_{n+1}}{a_n} < \rho + \varepsilon. \tag{8}$$

Since $\rho < 1$, we can take $\varepsilon$ so small that the inequality $\rho + \varepsilon < 1$ will be satisfied. Setting $\rho + \varepsilon = q$, we have, by virtue of the right-hand inequality (8),

$$\frac{a_{n+1}}{a_n} < q, \quad \text{or} \quad a_{n+1} < a_n q$$

* D'Alembert, Jean le Rond (1717-1783), a French mathematician, physicist and philosopher.

for $n = N, N+1, N+2, \ldots$. Assigning these values to $n$, we find from the last inequality that

$$a_{N+1} < a_N q,$$
$$a_{N+2} < a_{N+1} q < a_N q^2,$$
$$a_{N+3} < a_{N+2} q < a_N q^3,$$
$$\cdots\cdots\cdots\cdots$$

i.e. the terms of the series

$$a_{N+1} + a_{N+2} + a_{N+3} + \cdots \tag{9}$$

are smaller than the corresponding terms of the series composed of the elements of the geometric progression

$$a_N q + a_N q^2 + a_N q^3 + \cdots. \tag{10}$$

Since $q < 1$, series (10) is conevrgent (see Example 3 in 3.14.1). Then, according to the comparison test, series (9) is also convergent. But we have obtained series (9) from the given series $\sum_{n=1}^{\infty} a_n$ as a result of removal of a finite number of the first terms and, consequently, according to Theorem 14.1, the series $\sum_{n=1}^{\infty} a_n$ is also convergent.

(b) Let now $\rho > 1$. We shall prove that the series $\sum_{n=1}^{\infty} a_n$ is divergent. We take $\varepsilon$ so small that $\rho - \varepsilon > 1$. Then, for $n \geqslant N$, by virtue of the left-hand inequality (8), the inequality $\frac{a_{n+1}}{a_n} > 1$ is satisfied, or $a_{n+1} > a_n$. Thus, beginning with some number $N$, the terms of the series increase with an increase in their numbers, i.e. the general term $a_n$ of the series tends to zero as $n \to \infty$. Consequently, according to Theorem 14.4, the series $\sum_{n=1}^{\infty} a_n$ is divergent. ■

**Remark.** As can be seen from examples, for $\rho = 1$ the series $\sum_{n=1}^{\infty} a_n$ may be convergent or divergent. In this case an additional investigation of the series is needed with the aid of the comparison test or some other tests.

**Example 3.** The series $\sum_{n=1}^{\infty} \frac{1}{n!}$ is convergent since

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = \lim_{n\to\infty} \frac{n!}{(n+1)!} = \lim_{n\to\infty} \frac{1}{n+1} = 0 < 1.$$

**Example 4.** The series $\sum_{n=1}^{\infty} \frac{n^n}{n!}$ is divergent since

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = \lim_{n\to\infty} \frac{(n+1)^{n+1} n!}{(n+1)! n^n} = \lim_{n\to\infty} \left(\frac{n+1}{n}\right)^n = \lim_{n\to\infty} \left(1+\frac{1}{n}\right)^n = e > 1.$$

**Example 5.** Let us consider a series $\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$. We have $\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = \lim_{n\to\infty} \frac{\sqrt{n}}{\sqrt{n+1}} = 1$. D'Alembert's test does not allow us to infer whether the series is convergent or divergent. However, we have shown earlier (see Example 2) that the series is divergent.

**Integral test.**

**Theorem 14.8.** *Consider a series*

$$f(1) + f(2) + f(3) + \ldots + f(n) + \ldots = \sum_{n=1}^{\infty} f(n),$$

*whose terms are the values of a function $f(x)$ which is positive, continuous and decreasing on the half-open interval $[1, +\infty)$. If $\int_1^{+\infty} f(x)\,dx$ is convergent, then the series $\sum_{n=1}^{\infty} f(n)$ is also convergent, and if $\int_1^{+\infty} f(x)\,dx$ is divergent, then the series $\sum_{n=1}^{\infty} f(n)$ is also divergent.*

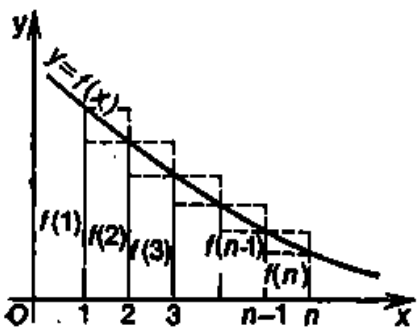



Fig. 214

*Proof.* We shall consider a curvilinear trapezoid bounded by the graph of the function $y = f(x)$ from above, by the straight lines $x = 1$ and $x = n$ from the nonparallel sides and by the $x$-axis from below. We inscribe into this trapezoid and circumscribe about it two stepped figures consisting of rectangles with bases $[1, 2], [2, 3], \ldots, [n-1, n]$ and altitudes $f(1), f(2), f(3), \ldots, f(n-1), f(n)$ (Fig. 214). Then, bearing in mind the geometric meaning of the definite integral, we have

$$f(2) + f(3) + \ldots + f(n)$$

$$< \int_1^n f(x)\,dx < f(1) + f(2) + \ldots + f(n-1),$$

or, in a concise form,

$$S_n - f(1) < \int_1^n f(x)\,dx < S_n - f(n).$$

From this we find that

$$S_n < f(1) + \int_1^n f(x)\,dx, \tag{11}$$

$$S_n > f(n) + \int_1^n f(x)\,dx, \tag{12}$$

where $S_n$ are the partial sums of the series in question.

Let the integral $\int_1^{+\infty} f(x)\,dx$ be convergent. This means that there is a limit $\lim\limits_{n\to\infty} \int_1^n f(x)\,dx = I$. Since $f(x) > 0$, the sequence $\int_1^n f(x)\,dx$ increases with an increase in $n$ and is bounded by its limit from above: $\int_1^n f(x)\,dx < I$. It follows from inequality (11) that $S_n < f(1) + I$, i.e. the sequence $\{S_n\}$ of partial sums of the series $\sum\limits_{n=1}^{\infty} f(n)$ is bounded. According to Theorem 14.5, the series $\sum\limits_{n=1}^{\infty} f(n)$ is convergent.

Let now the integral $\int_1^{+\infty} f(x)\,dx$ be divergent. In this case $\int_1^n f(x)\,dx \to +\infty$ as $n \to \infty$ (being a monotone increasing unbounded sequence). It follows from inequality (12) that $S_n \to +\infty$ as $n \to \infty$, i.e. the sequence $\{S_n\}$ of partial sums of the series $\sum\limits_{n=1}^{\infty} f(n)$ is divergent and, consequently, the series is divergent. ■

**Example 6.** Let us consider a series $1 + \frac{1}{2^\alpha} + \frac{1}{3^\alpha} + \ldots + \frac{1}{n^\alpha} + \ldots = \sum\limits_{n=1}^{\infty} \frac{1}{n^\alpha}$ $(\alpha > 0)$. Using the integral test, we shall investigate the behaviour of this series for $\alpha > 0$. We take a function

$\frac{1}{x^\alpha}$ $(x \geqslant 1)$ as the function $f(x)$, the former satisfying the conditions of Theorem 14.8. The terms of the series are equal to the values of this function for $x = 1, 2, 3, \ldots, n, \ldots$. As is known (see Example 4 in 1.8.11, Item 1) the improper integral $\int\limits_1^{+\infty} \frac{dx}{x^\alpha}$ is convergent for $\alpha > 1$ and divergent for $\alpha \leqslant 1$. Consequently, this series is convergent for $\alpha > 1$ and divergent for $\alpha \leqslant 1$.

Note that for $\alpha \leqslant 0$ series of this kind are also divergent since their general term tends to zero as $n \to \infty$, i.e. the necessary condition for convergence of a series is violated (see Theorem 14.4).

In particular, for $\alpha = 2$ we have a convergent series $\sum\limits_{n=1}^{\infty} \frac{1}{n^2}$, for $\alpha = 1$ we have a divergent harmonic series $\sum\limits_{n=1}^{\infty} \frac{1}{n}$, for $\alpha = \frac{1}{2}$ we have a divergent series $\sum\limits_{n=1}^{\infty} \frac{1}{\sqrt{n}}$, and so on.

### 3.14.3. Alternating Series

Up till now we considered series with nonnegative terms. Series with nonpositive terms differ from the respective series with nonnegative terms only by a factor $-1$ and therefore the tests for their convergence are similar.

We shall consider now alternating series whose terms have alternating signs. For the sake of convenience, we assume that the first term of such a series is positive. Then we can write an alternating series as

$$a_1 - a_2 + a_3 - a_4 + \ldots + (-1)^{n+1} a_n + \ldots, \qquad (1)$$

where $a_n > 0$.

The following simple sufficient test for convergence holds true for alternating series.

**Leibniz' test.**

**Theorem 14.9.** *If the absolute values of the terms of the alternating series* (1) *decrease monotonically, i.e.* $a_1 > a_2 > a_3 > \ldots$ *and the general term of the series tends to zero, i.e.* $\lim\limits_{n\to\infty} a_n = 0$, *then the series is convergent.*

*Proof.* Given series (1), let $a_n > a_{n+1}$ and $a_n \to 0$ as $n \to \infty$.

We shall consider a partial sum of the series with an even number of terms

$$S_{2n} = a_1 - a_2 + a_3 - a_4 + \ldots + a_{2n-1} - a_{2n}$$
$$= (a_1 - a_2) + (a_3 - a_4) + \ldots + (a_{2n-1} - a_{2n}).$$

By virtue of the first condition, all the differences in parentheses are positive, and therefore the sequence $\{S_{2n}\}$ of partial sums is increasing. We shall prove that it is bounded. To do this, we represent $S_{2n}$ in the form

$$S_{2n} = a_1 - [(a_2 - a_3) + (a_4 - a_5) + \ldots + (a_{2n-2} - a_{2n-1}) + a_{2n}].$$

It follows that $S_{2n} < a_1$ for any $n$, i.e. $\{S_{2n}\}$ is bounded.

Thus the sequence $\{S_{2n}\}$ is increasing, bounded and, consequently, has a limit $\lim\limits_{n\to\infty} S_{2n} = S$.

We shall show now that the sequence of partial sums of an odd number of terms also converges to the same limit $S$. Indeed, $S_{2n+1} = S_{2n} + a_{2n+1}$. Passing to the limit in this relation as $n \to \infty$ and using the second condition ($a_n \to 0$ as $n \to \infty$), we obtain

$$\lim_{n\to\infty} S_{2n+1} = \lim_{n\to\infty} (S_{2n} + a_{2n+1}) = \lim_{n\to\infty} S_{2n} + \lim_{n\to\infty} a_{2n+1} = S + 0 = S.$$

Thus the sequence $\{S_n\}$ of partial sums of series (1) converges to the limit $S$. And this means that series (1) is convergent. ■

**Example.** The series $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots + (-1)^{n+1}\frac{1}{n} + \ldots = \sum\limits_{n=1}^{\infty} (-1)^{n+1}\frac{1}{n}$ is convergent since it satisfies the conditions of Leibniz' test: (1) $1 > \frac{1}{2} > \frac{1}{3} > \ldots$; (2) $\lim\limits_{n\to\infty} \frac{1}{n} = 0$. Note that this series differs from the harmonic series only by the signs of the even terms.

### 3.14.4. Absolute and Conditional Convergence of Series

Consider now series with terms of arbitrary signs.

We take a series

$$a_1 + a_2 + a_3 + \ldots + a_n + \ldots = \sum_{n=1}^{\infty} a_n, \tag{1}$$

where the numbers $a_1, a_2, a_3, \ldots, a_n, \ldots$ may be positive or negative, the positions of positive and negative terms in the series being arbitrary. At the same time we consider a series composed of the

absolute values of the terms of series (1):

$$|a_1| + |a_2| + |a_3| + \ldots + |a_n| + \ldots = \sum_{n=1}^{\infty} |a_n|. \tag{2}$$

The following test for convergence holds true for series of this kind.

**Theorem 14.10.** *If series* (2) *is convergent, then series* (1) *is convergent too.*

*Proof.* Let series (2) be convergent. We denote the partial sum of series (1) by $S_n$ and the partial sum of series (2) by $\sigma_n$: $S_n = a_1 + a_2 + a_3 + \ldots + a_n$, $\sigma_n = |a_1| + |a_2| + |a_3| + \ldots + |a_n|$. Since series (2) is convergent, the sequence $\{\sigma_n\}$ of its partial sums has a limit $\lim\limits_{n\to\infty} \sigma_n = \sigma$, and the inequality

$$\sigma_n \leqslant \sigma \tag{3}$$

holds true for any $n$ since the terms of series (2) are nonnegative.

We denote the sum of positive terms by $S'_n$ and the sum of the absolute values of negative terms, which appear in the sum $S_n$, by $S''_n$. Then

$$S_n = S'_n - S''_n, \tag{4}$$

$$\sigma_n = S'_n + S''_n. \tag{5}$$

The sequences $\{S'_n\}$ and $\{S''_n\}$ are evidently nondecreasing and from equality (5) and inequality (3) it follows that they are bounded: $S'_n \leqslant \sigma_n \leqslant \sigma$ and $S''_n \leqslant \sigma_n \leqslant \sigma$. Consequently, there are limits $\lim\limits_{n\to\infty} S'_n = S'$ and $\lim\limits_{n\to\infty} S''_n = S''$. But in this case, by virtue of equality (4), the sequence of partial sums of series (1) has a limit

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty} (S'_n - S''_n) = \lim_{n\to\infty} S'_n - \lim_{n\to\infty} S''_n = S' - S''.$$

And this means that series (1) is convergent. ■

**Example 1.** The series $1 - 1/2^2 - 1/3^2 + 1/4^2 + 1/5^2 - 1/6^2 - 1/7^2 + \ldots$ is convergent according to the test we have proved since the series composed of the absolute values of the terms of this series is convergent: $1 + 1/2^2 + 1/3^2 + 1/4^2 + 1/5^2 + 1/6^2 + 1/7^2 + \ldots$ (see Example 6 in 3.14.2).

The test for the convergence of a series with arbitrary signs we have just considered is sufficient but not necessary since there are series of this kind which are convergent, whereas the series composed of the absolute values of their terms are divergent. Thus, for instance, the series $\sum\limits_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n}$ is convergent according to Leibniz' test (see the example in 3.14.3), whereas the

series $\sum_{n=1}^{\infty} \frac{1}{n}$ composed of the absolute values of its terms is divergent (a harmonic series).

Therefore we can classify all convergent series into absolutely convergent and conditionally convergent series.

*Absolutely convergent series* are series for which the series composed of the absolute values of their terms are also convergent.

*Conditionally convergent series* are series for which the series composed of the absolute values of their terms are divergent.

**Example 2.** The series $1 - 1/2 + 1/4 - 1/8 + 1/16 - 1/32 + \ldots$ is absolutely convergent since the series composed of the absolute values, $1 + 1/2 + 1/4 + 1/8 + 1/16 + 1/32 + \ldots$, is also convergent. (Both series are geometric progressions with common ratios equal to $-1/2$ and $1/2$ respectively.)

**Example 3.** The series $1 - 1/\sqrt{2} + 1/\sqrt{3} - 1/\sqrt{4} + \ldots$ is conditionally convergent since it itself is convergent according to Leibniz' test, but the series composed of the absolute values, $1 + 1/\sqrt{2} + 1/\sqrt{3} + 1/\sqrt{4} + \ldots$, is divergent (see Example 6 in 3.14.2).

Note that the division of convergent series into absolutely and conditionally convergent is essential since absolutely convergent series possess a number of important properties, whereas conditionally convergent series do not possess some of those properties. For instance, for conditionally convergent series the sum of the series is not equal to the sum of positive and the sum of negative terms of the series as is the case with absolutely convergent series. We have shown this when proving Theorem 14.10.

### 3.14.5. Power Series

**1. Definition and general remarks.** *A series of the form*

$$a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots + a_nx^n + \ldots = \sum_{n=0}^{\infty} a_nx^n \qquad (1)$$

*is a power series.*

The numbers $a_0, a_1, a_2, a_3, \ldots, a_n, \ldots$ are the *coefficients of the power series.*

Assigning various numerical values to $x$, we obtain different power series which may be convergent or divergent. The set of the values of $x$ for which series (1) is convergent is the *domain of its convergence.* This set is always nonempty since any power series converges for $x = 0$.

The partial sum of the power series $S_n(x) = a_0 + a_1x + \ldots + a_nx^n$ is evidently a function of the variable $x$. Therefore the sum $S$

is also a function of the variable $x$ defined in the domain of convergence of the series: $S=S(x)=\sum_{n=0}^{\infty} a_n x^n$ $\left(\text{or } f(x)=\sum_{n=0}^{\infty} a_n x^n\right)$.

**2. The interval of convergence of a power series.** We shall prove a theorem which is significant in the theory of power series and deals with the domain of convergence of a power series.

**Theorem 14.11 (Abel's* theorem on power series).** (1) *If the power series* (1) *is convergent for* $x = x_0$ $(x_0 \neq 0)$, *then it is convergent for all* $x$ *satisfying the condition* $|x| < |x_0|$, *the convergence being absolute*; (2) *if series* (1) *is divergent for* $x = x_1$, *then it is divergent for all* $x$ *satisfying the condition* $|x| > |x_1|$.

*Proof.* (1) Since by the hypothesis the numerical series $\sum_{n=0}^{\infty} a_n x_0^n$ is convergent, its general term $a_n x_0^n \to 0$ as $n \to \infty$, whence it follows that the sequence $\{a_n x_0^n\}$ is bounded, i.e. there is a number $M > 0$ such that

$$|a_n x_0^n| < M, \; n = 0, 1, 2, \ldots . \tag{2}$$

We rewrite series (1) in the form

$$a_0 + a_1 x_0 \left(\frac{x}{x_0}\right) + a_2 x_0^2 \left(\frac{x}{x_0}\right)^2 + \ldots + a_n x_0^n \left(\frac{x}{x_0}\right)^n + \ldots \tag{3}$$

and consider the series composed of the absolute values of its terms

$$|a_0| + |a_1 x_0| \left|\frac{x}{x_0}\right| + |a_2 x_0^2| \left|\frac{x}{x_0}\right|^2 + \ldots + |a_n x_0^n| \left|\frac{x}{x_0}\right|^n + \ldots . \tag{4}$$

By virtue of inequality (2), the terms of series (4) are smaller than the corresponding terms of the series

$$M + M\left|\frac{x}{x_0}\right| + M\left|\frac{x}{x_0}\right|^2 + \ldots + M\left|\frac{x}{x_0}\right|^n + \ldots . \tag{5}$$

For $|x| < |x_0|$ series (5) is a geometric progression with the common ratio $q = \left|\frac{x}{x_0}\right| < 1$ and, consequently, is convergent. Since the terms of series (4) are smaller than the corresponding terms of series (5), series (4) is also convergent according to the comparison test, and this means that for $|x| < |x_0|$ series (1) is absolutely convergent.

(2) We shall prove now the second part of the theorem. By the hypothesis, series (1) is divergent at the point $x_1$. We have to prove that it is divergent for all $x$ which satisfy the condition $|x| > |x_1|$. Let us assume the contrary, i.e. that series (1) is convergent for a certain value of $x$ such that $|x| > |x_1|$. Then, in accordance with what we have just proved for the first part of the theorem,

* Abel, Niels Henrik (1802-1829), a Norwegian mathematician.

series (1) must converge at the point $x_1$ as well since $|x_1| < |x|$. But this contradicts the fact that the series diverges at the point $x_1$. ■

Abel's theorem states that if $x_0$ is a point of convergence of a power series, then the series is absolutely convergent at all points belonging to the interval $(-|x_0|, |x_0|)$ (Fig. 215*a*), and if $x_1$ is a

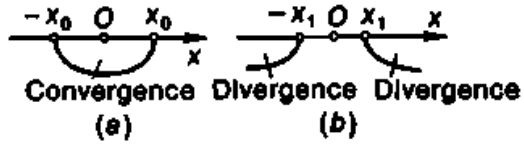



Fig. 215

point of divergence of a power series, then the series is divergent at all points which do not belong to the interval $(-|x_1|, |x_1|)$ (Fig. 215*b*).

Hence follows a theorem.

**Theorem 14.12.** *If the series* $\sum_{n=0}^{\infty} a_n x^n$ *is convergent not for all values of* $x$ *and not only for* $x = 0$, *then there is a number* $R > 0$ *such that the series is absolutely convergent for* $|x| < R$ *and divergent for* $|x| > R$.

*Proof.* We designate the set of points $x$ at which the series $\sum_{n=0}^{\infty} a_n x^n$ is convergent as $X$. We shall show that the set $X$ is bounded. Indeed, if we take a point $x_1$ at which the series is divergent (points of this kind exist by the hypothesis), then, according to Abel's theorem, the inequality $|x| < |x_1|$ is satisfied for any $x$ from $X$. A set bounded from above is known to have the least upper bound. We set $R = \sup_{x \in X} |x|$. Since the series is convergent not only for $x = 0$, it follows that $R > 0$.

We shall take now any $x$ for which $|x| < R$. According to the property of the least upper bound, there is $x_0 \in X$ such that $|x| < |x_0| \leqslant R$, whence, in accordance with Abel's theorem, there follows the absolute convergence of the series for the $x$ we have taken.

Let us take now any $x$ for which $|x| > R$ such that $x \notin X$. Consequently, for this $x$ the series is divergent. ■

We have thus solved the problem concerning the domain of convergence of a power series. The interval $(-R, R)$ is the *interval of convergence of a power series.* The number $R$ is the *radius of convergence of a power series.*

Note that the interval of convergence of some series is the entire number line (in this case we write $R = \infty$), whereas for other series the interval can degenerate into a single point ($R = 0$).

Thus every power series has its own radius of convergence $R$.

For $x = \pm R$, a series may either converge or diverge, the case being specific for every definite series.

Here is a method of determining the radius of convergence of a power series.

**Theorem 14.13.** *If there is a limit* $\lim\limits_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| \neq 0$, *then the radius of convergence of the series* $\sum\limits_{n=0}^{\infty} a_n x^n$ *is* $R = \lim\limits_{n\to\infty}\left|\frac{a_n}{a_{n+1}}\right|$.

*Proof.* Let us consider the series $\sum\limits_{n=0}^{\infty} |a_n x^n|$. By the hypothesis, there is a limit $\lim\limits_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| \neq 0$. We designate it as $\frac{1}{R}$. Then

$$\lim_{n\to\infty}\left|\frac{a_{n+1}x^{n+1}}{a_n x^n}\right| = |x| \lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = \frac{|x|}{R}.$$

For every value of $x$ the power series becomes a number series. Therefore, according to d'Alembert's test, the series $\sum\limits_{n=0}^{\infty} |a_n x^n|$ is convergent if $\frac{|x|}{R} < 1$, i.e. $|x| < R$.

Consequently, in accordance with Theorem 14.10 on the convergence of series with arbitrary signs, the series $\sum\limits_{n=0}^{\infty} a_n x^n$ is also convergent for $|x| < R$, the convergence being absolute. For $|x| > R$ the series $\sum\limits_{n=0}^{\infty} a_n x^n$ is divergent since $\lim\limits_{n\to\infty}\left|\frac{a_{n+1}x^{n+1}}{a_n x^n}\right| = \frac{|x|}{R} > 1$ and, consequently, the general term $a_n x^n$ of the series does not tend to zero as $n \to \infty$.

Thus this series is convergent inside the interval $(-R, R)$ and divergent outside it, i.e. the radius of convergence $R = \lim\limits_{n\to\infty}\left|\frac{a_n}{a_{n+1}}\right|$. ■

**Remark.** We can prove that if $\lim\limits_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = 0$, then the series $\sum\limits_{n=0}^{\infty} a_n x^n$ converges throughout the number line, i.e. $R = \infty$, and if $\lim\limits_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = \infty$, then the series converges only for $x = 0$, i.e. $R = 0$.

**Example 1.** Consider a series $\sum_{n=1}^{\infty} \frac{|x^n|}{n}$. Here $a_n = \frac{1}{n}$ and $a_{n+1} = \frac{1}{n+1}$. Therefore

$$R = \lim_{n\to\infty} \frac{a_n}{a_{n+1}} = \lim_{n\to\infty} \frac{n+1}{n} = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right) = 1.$$

Consequently, according to Theorem 14.13, this series converges on the interval $(-1, 1)$. We shall investigate the behaviour of the series at the endpoints of the interval of convergence, i.e. at the points $x = \pm 1$. For $x = 1$ we obtain a harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$ and for $x = -1$ a series $\sum_{n=1}^{\infty} (-1)^n \frac{1}{n}$, which is convergent by virtue of Leibniz' test. Thus this series is convergent at any point of the half-open interval $[-1, 1)$ and divergent outside it.

**Example 2.** The series $\sum_{n=1}^{\infty} n! x^n$ is divergent throughout the number line, except for the point $x = 0$, since the radius of convergence

$$R = \lim_{n\to\infty} \frac{a_n}{a_{n+1}} = \lim_{n\to\infty} \frac{n!}{(n+1)!} = \lim_{n\to\infty} \frac{1}{1+n} = 0.$$

**Example 3.** The series $\sum_{n=1}^{\infty} \frac{x^n}{n!}$ is absolutely convergent throughout the number line since its radius of convergence

$$R = \lim_{n\to\infty} \frac{a_n}{a_{n+1}} = \lim_{n\to\infty} \frac{(n+1)!}{n!} = \lim_{n\to\infty} (n+1) = \infty.$$

**3. Properties of power series.** Let the function $f(x)$ be the sum of the power series

$$f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n + \ldots \tag{6}$$

whose interval of convergence is $(-R, R)$.

In this case we say that on the interval $(-R, R)$ the function $f(x)$ can be expanded into a power series (or a series in the powers of $x$).

There are two theorems on the properties of power series which will be given here without a proof.

**Theorem 14.14.** *If, on the interval $(-R, R)$, the function $f(x)$ can be expanded into the power series* (6), *then it is differentiable on this interval and its derivative $f'(x)$ can be found by means of a term-by-*

*term differentiation of series* (6), *i.e.*

$$f'(x) = (a_0 + a_1x + a_2x^2 + \ldots + a_nx^n + \ldots)'$$
$$= a_1 + 2a_2x + 3a_3x^2 + \ldots + na_nx^{n-1} + \ldots .$$

*We can similarly calculate derivatives of any order of the function* $f(x)$. *In this case the corresponding series have the same interval of convergence as series* (6).

**Theorem 14.15.** *If, on the interval* $(-R, R)$, *the function* $f(x)$ *can be expanded into the power series* (6), *then it is integrable in the interval* $(-R, R)$ *and the integral of this function can be calculated by means of a term-by-term integration of series* (6), *i.e. if* $x_1, x_2 \in (-R, R)$, *then*

$$\int_{x_1}^{x_2} f(x)\,dx = \int_{x_1}^{x_2} (a_0 + a_1x + a_2x^2 + \ldots + a_nx^n + \ldots)\,dx$$
$$= \int_{x_1}^{x_2} a_0\,dx + \int_{x_1}^{x_2} a_1x\,dx + \int_{x_1}^{x_2} a_2x^2\,dx + \int_{x_1}^{x_2} a_nx^n\,dx + \ldots .$$

Also of interest is the integration of the power series (6) over the interval $[0, x]$, where $|x| < R$:

$$\int_0^x f(x)\,dx = a_0x + \frac{a_1x^2}{2} + \frac{a_2x^3}{3} + \ldots + \frac{a_nx^{n+1}}{n+1} + \ldots .$$

In this case we again get a power series which has the same interval of convergence as series (6).

The theorems on the differentiation and integration of power series which we have formulated are very important and will be repeatedly used in what follows.

Note that in a number of cases we consider power series of a more general form:

$$a_0 + a_1(x-a) + a_2(x-a)^2 + \ldots + a_n(x-a)^n + \ldots$$
$$= \sum_{n=0}^{\infty} a_n(x-a)^n. \tag{7}$$

A series of form (7) can be reduced to form (1) by a change of variable $x - a = t$.

If the function $f(x)$ is the sum of series (7), then we say that the function $f(x)$ can be expanded into a series in the powers of $(x - a)$.

Everything we have presented here can be extended to a series of form (7). To make the notation simpler, we present the arguments that follow for series of form (1).

**4. Expansion of functions into power series.** As is shown by the following theorem, the expansion of a function into a power series is unique.

**Theorem 14.16.** *If, on the interval* $(-R, R)$, *the function* $f(x)$ *can be expanded into a power series*

$$f(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n + \ldots, \tag{8}$$

*then the expansion is unique.*

*Proof.* By the hypothesis, series (8) is convergent on the interval $(-R, R)$ and the function $f(x)$ is its sum. Consequently, on the basis of Theorem 14.14, series (8) can be differentiated term by term on the interval $(-R, R)$ any number of times. As a result of the differentiation, we get

$$\begin{aligned}
f'(x) &= 1\cdot a_1 + 2a_2x + 3a_3x^2 + 4a_4x^3 + \ldots + na_nx^{n-1} + \ldots,\\
f''(x) &= 1\cdot 2a_2 + 2\cdot 3a_3x + 3\cdot 4a_4x^2 + 4\cdot 5a_5x^3 + \ldots\\
&\quad + n(n-1)a_nx^{n-2} + \ldots,\\
f'''(x) &= 1\cdot 2\cdot 3a_3 + 2\cdot 3\cdot 4a_4x + \ldots + n(n-1)(n-2)a_nx^{n-3} + \ldots,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
f^{(n)}(x) &= n!a_n + 2\cdot 3\cdot 4\ldots(n+1)a_{n+1}x + \ldots,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned}$$

Setting $x = 0$ in the relations obtained and in relation (8), we have

$$f(0) = a_0,\ f'(0) = 1\cdot a_1,\ f''(0) = 2!a_2,$$
$$f'''(0) = 3!a_3,\ \ldots,\ f^{(n)}(0) = n!a_n,\ \ldots,$$

whence we find that

$$a_0 = f(0),\ a_1 = \frac{f'(0)}{1!},\ a_2 = \frac{f''(0)}{2!},$$
$$a_3 = \frac{f'''(0)}{3!},\ \ldots,\ a_n = \frac{f^{(n)}(0)}{n!},\ \ldots. \tag{9}$$

Thus all the coefficients of series (8) are uniquely defined by formulas (9), and this proves the theorem. ■

Substituting the expressions obtained for the coefficients into relation (8), we get

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \ldots + \frac{f^{(n)}(0)}{n!}x^n + \ldots.$$

Thus, if the function $f(x)$ can be expanded into a power series, then the series has the form

$$f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \ldots + \frac{f^{(n)}(0)}{n!}x^n + \ldots. \tag{10}$$

Series (10) is known as *Maclaurin's series* for the function $f(x)$.

Let now $f(x)$ be an arbitrary infinitely differentiable function. We can compose series (10) for it. Let us find the conditions under which the sum of series (10) coincides with the function $f(x)$. We can get the answer to this question by means of Maclaurin's formula. As was shown in 1.6.3, for any infinitely differentiable function Maclaurin's formula

$$f(x)=f(0)+\frac{f'(0)}{1!}x+\frac{f''(0)}{2!}x^2+\ldots+\frac{f^{(n)}(0)}{n!}x^n+R_n(x)$$

holds true, where the remainder

$$R_n(x)=\frac{f^{(n+1)}(\xi)}{(n+1)!}x^{n+1},\quad \xi=\theta x,\quad 0<\theta<1. \tag{11}$$

If we designate the partial sum of Maclaurin's series as $S_n(x)$, we can write Maclaurin's formula as

$$f(x)=S_n(x)+R_n(x). \tag{12}$$

Consider the following theorem.

**Theorem 14.17.** *For Maclaurin's series* (10) *to be convergent on* $(-R, R)$ *and to have the function* $f(x)$ *as its sum, it is necessary and sufficient that on* $(-R, R)$ *the remainder* $R_n(x)$ *of Maclaurin's formula* (11) *tend to zero as* $n\to\infty$, *i.e.* $\lim\limits_{n\to\infty} R_n(x)=0$ *for any* $x\in(-R, R)$.

*Proof.* **Necessity.** Let the function $f(x)$ be the sum of Maclaurin's series on $(-R, R)$, i.e. $\lim\limits_{n\to\infty} S_n(x)=f(x)$. Then it follows from relation (12) that $\lim\limits_{n\to\infty} R_n(x)=0$ for any $x\in(-R, R)$.

**Sufficiency.** Let $\lim\limits_{n\to\infty} R_n(x)=0$ for any $x\in(-R, R)$. Then it follows from relation (12) that $\lim\limits_{n\to\infty}[f(x)-S_n(x)]=0$, i.e. $\lim\limits_{n\to\infty} S_n(x)=f(x)$. And this means that Maclaurin's series (10) is convergent on $(-R, R)$ and its sum is equal to $f(x)$. ■

It follows from the theorem that the problem of expanding a function into a Maclaurin's series reduces to the investigation of the behaviour of the remainder $R_n(x)$ as $n\to\infty$.

Consider the expansions of some elementary functions into a Maclaurin's series.

**Expansion of the function** $f(x)=e^x$. We have $f'(x)=f''(x)=\ldots=f^{(n)}(x)=e^x$, whence, for $x=0$, we get $f(0)=f'(0)=f''(0)=\ldots=f^{(n)}(0)=1$. From formula (10) for the function $e^x$ we compose a Maclaurin's series

$$1+\frac{x}{1!}+\frac{x^2}{2!}+\ldots+\frac{x^n}{n!}+\ldots. \tag{13}$$

Then we find the interval of convergence of series (13):

$$R=\lim_{n\to\infty}\frac{a_n}{a_{n+1}}=\lim_{n\to\infty}\frac{n!(n+1)}{n!}=\infty.$$

Consequently, the series is absolutely convergent throughout the number line.

We shall prove now that the function $e^x$ is the sum of series (13).

Note that, by virtue of the necessary condition for the convergence of a series for any $x$, there holds an equality

$$\lim_{n\to\infty}\frac{|x|^n}{n!}=0. \tag{14}$$

Since $f^{(n+1)}(\xi)=e^{\xi}$, it follows that

$$R_n(x)=\frac{f^{(n+1)}(\xi)}{(n+1)!}x^{n+1}=\frac{e^{\xi}}{(n+1)!}x^{n+1},$$

where $\xi=\theta x$, $0<\theta<1$. From this, taking into account that $e^{\xi}<e^{|x|}$, we obtain

$$|R_n(x)|=\frac{e^{\xi}}{(n+1)!}|x|^{n+1}<\frac{e^{|x|}}{(n+1)!}|x|^{n+1}.$$

Since

$$\lim_{n\to\infty}\frac{|x|^n}{n!}=0$$

it follows, by virtue of (14), that

$$\lim_{n\to\infty}\frac{|x|^{n+1}}{(n+1)!}=0$$

as well. Therefore, passing to the limit in the last inequality as $n\to\infty$, we find that $\lim\limits_{n\to\infty} R_n(x)=0$ for any $x$ and, consequently, the function $e^x$ is the sum of series (13).

Thus, for any $x$ there holds an expansion

$$e^x=1+\frac{x}{1!}+\frac{x^2}{2!}+\ldots+\frac{x^n}{n!}+\ldots.$$

**Expansion of the function** $f(x)=\sin x$. We have $f'(x)=\cos x=\sin\left(x+\frac{\pi}{2}\right)$, $f''(x)=-\sin x=\sin\left(x+2\frac{\pi}{2}\right)$, ..., $f^n(x)=\sin\left(x+n\frac{\pi}{2}\right)$ (see 1.5.10, Item 2), whence, setting $x=0$, we find that $f(0)=0$, $f'(0)=1$, $f''(0)=0$, $f'''(0)=-1$, $f^{(4)}(0)=0$, .... Let us compose a Maclaurin's series

$$x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\ldots+\frac{(-1)^{n-1}x^{2n-1}}{(2n-1)!}+\ldots$$

from formula (10) for the function $\sin x$.

It is easy to verify that the series obtained is absolutely convergent throughout the number line. Let us investigate the remainder

$$R_n(x)=\frac{f^{(n+1)}(\xi)}{(n+1)!}x^{n+1}=\frac{\sin\left[\xi+(n+1)\frac{\pi}{2}\right]}{(n+1)!}x^{n+1},$$

where $\xi=\theta x$, $0<\theta<1$. Since $\left|\sin\left[\xi+(n+1)\frac{\pi}{2}\right]\right|\leqslant 1$, it follows that $|R_n(x)|\leqslant\frac{|x|^{n+1}}{(n+1)!}$. By virtue of formula (14), $\lim\limits_{n\to\infty}\frac{|x|^{n+1}}{(n+1)!}=0$. Consequently, $\lim\limits_{n\to\infty}R_n(x)=0$ for any $x$. And this means that the function $\sin x$ is the sum of the series we have constructed, i.e. there holds an expansion

$$\sin x=x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\ldots+\frac{(-1)^{n-1}x^{2n-1}}{(2n-1)!}+\ldots.$$

**Expansion of the function $f(x)=\cos x$.** By analogy with what we did, we can get the expansion of the function $\cos x$ into a Maclaurin's series valid for any $x$. However, it is even easier to obtain the expansion of $\cos x$ by means of a term-by-term differentiation of the series for $\sin x$:

$$(\sin x)'=(x)'-\left(\frac{x^3}{3!}\right)'+\left(\frac{x^5}{5!}\right)'-\left(\frac{x^7}{7!}\right)'+\ldots$$
$$+\left[\frac{(-1)^{n-1}x^{2n-1}}{(2n-1)!}\right]'+\ldots,$$

whence

$$\cos x=1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\ldots+(-1)^n\frac{x^{2n}}{(2n)!}+\ldots.$$

In addition to the functions $e^x$, $\sin x$, $\cos x$, many other functions can be expanded into a Maclaurin's series. Instead of a Maclaurin's series we could consider a more general Taylor's series in the powers of $(x-a)$, where $a\neq 0$, i.e. a series of the form

$$f(a)+\frac{f'(a)}{1!}(x-a)+\frac{f''(a)}{2!}(x-a)^2+\ldots+\frac{f^{(n)}(a)}{n!}(x-a)^n+\ldots.$$

All we have presented can be completely extended to series of this kind.

When expanding the function $\cos x$ into a Maclaurin's series, we have made use of the property of a termwise differentiation of power series. We can also use another property of power series, their termwise integrability. By way of example, we shall expand the functions $\ln(1+x)$ and $\arctan x$ into power series using a termwise integration.

Let us consider a series $1 + x + x^2 + x^3 + \ldots + x^n + \ldots$. This is a geometric progression whose first term is unity and the common ratio $q = x$. As is known, this series converges for $|x| < 1$ and its sum is $1/(1 - x)$. Consequently,

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots + x^n + \ldots. \tag{15}$$

Relation (15) is an expansion of the function $f(x) = 1/(1 - x)$ into a power series.

Substituting $-t$ for $x$ into relation (15), we get a relation

$$\frac{1}{1+t} = 1 - t + t^2 - t^3 + \ldots + (-1)^n t^n + \ldots,$$

which is valid for $|t| < 1$. We perform a termwise integration of this power series in the limits from 0 to $x$ ($|x| < 1$) and obtain

$$\int_0^x \frac{dt}{1+t} = \ln(1+t)\Big|_0^x = \ln(1+x)$$

$$= \int_0^x [1 - t + t^2 - t^3 + \ldots + (-1)^n t^n + \ldots]\, dt$$

$$= \int_0^x dt - \int_0^x t\, dt + \int_0^x t^2\, dt - \int_0^x t^3\, dt + \ldots + (-1)^n \int_0^x t^n\, dt + \ldots$$

$$= t\Big|_0^x - \frac{t^2}{2}\Big|_0^x + \frac{t^3}{3}\Big|_0^x - \frac{t^4}{4}\Big|_0^x + \ldots + (-1)^n \frac{t^{n+1}}{n+1}\Big|_0^x + \ldots$$

$$= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots + (-1)^n \frac{x^{n+1}}{n+1} + \ldots,$$

whence it follows that

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots + (-1)^n \frac{x^{n+1}}{n+1} + \ldots. \tag{16}$$

Relation (16) is the expansion of the function $\ln(1 + x)$ into a power series. It is valid for $|x| < 1$. We can prove that this equality is also valid for $x = 1$.

Indeed, for $x = 1$ the left-hand side of (16) is equal to $\ln 2$ and the right-hand side is a power series

$$1 - \frac{1}{2} + \frac{1}{3} - \ldots + (-1)^{n-1} \frac{1}{n} + \ldots \tag{17}$$

convergent according to Leibniz' test. It remains to verify the validity of the equality

$$\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \ldots + (-1)^{n-1} \frac{1}{n} + \ldots. \tag{18}$$

To do this, we integrate the expression

$$\frac{1}{1+t}=1-t+t^2-t^3+\ldots+(-1)^{n-1}t^{n-1}+(-1)^n\frac{t^n}{1+t},$$

obtained as a result of the division of unity by $1+t$, from 0 to 1. We have

$$\int_0^1 \frac{dt}{1+t}=\ln(1+t)\Big|_0^1=\ln 2$$

$$=\int_0^1\left[1-t+t^2-t^3+\ldots+(-1)^{n-1}t^{n-1}+(-1)^n\frac{t^n}{1+t}\right]dt$$

$$=\int_0^1 dt-\int_0^1 t\,dt+\int_0^1 t^2\,dt-\int_0^1 t^3\,dt+\ldots+(-1)^{n-1}\int_0^1 t^{n-1}\,dt$$

$$+(-1)^n\int_0^1\frac{t^n}{1+t}\,dt=1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\ldots+(-1)^{n-1}\frac{1}{n}$$

$$+(-1)^n\int_0^1\frac{t^n}{1+t}\,dt,$$

i.e.

$$\ln 2=1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\ldots+(-1)^{n-1}\frac{1}{n}$$

$$+(-1)^n\int_0^1\frac{t^n}{1+t}\,dt. \tag{19}$$

In this equality, the sum of the first $n$ terms is a partial sum $S_n$ of series (17). We write (19) in the form

$$\ln 2-S_n=(-1)^n\int_0^1\frac{t^n\,dt}{1+t}. \tag{20}$$

Since $t^n/(1+t)\leqslant t^n$ for $0\leqslant t\leqslant 1$, we have

$$(-1)^n\int_0^1\frac{t^n\,dt}{1+t}\Big|=\int_0^1\frac{t^n\,dt}{1+t}\leqslant\int_0^1 t^n\,dt=\frac{t^{n+1}}{n+1}\Big|_0^1=\frac{1}{n+1}\to 0 \text{ as } n\to\infty.$$

From this we infer that the integral on the right-hand side of (20) tends to zero as $n\to\infty$ and, consequently, $\lim\limits_{n\to\infty} S_n=\ln 2$, and this signifies the validity of equality (18).

We shall find now the expansion of the function arctan $x$. Substituting $-t^2$ for $x$ into (15) and integrating with respect to $t$ from 0 to $x$, we have

$$\arctan x = x - \frac{x^3}{3} + \frac{x^5}{5} - \ldots + (-1)^n \frac{x^{2n+1}}{2n+1} + \ldots . \qquad (21)$$

Equality (21) is valid for $|x| < 1$. We can prove by analogy with what we have proved earlier that it is also valid for $x = \pm 1$.

It must be noted in conclusion that power series have various applications. Using them, we can calculate the values of functions (in particular, the values of $\pi$ and e) with any degree of accuracy, find the approximate values of definite integrals which either cannot be expressed in terms of elementary functions or are difficult to calculate. Thus, for instance, the integral $\int_0^a \frac{\sin x}{x}\,dx$ cannot be taken with respect to elementary functions since the antiderivative of the function $(\sin x)/x$ is not elementary. At the same time this antiderivative can be easily expressed as a power series. Indeed, since $\sin x = x - x^3/3! + x^5/5! - x^7/7! + \ldots$, we can multiply this series by $1/x$ and obtain

$$\frac{\sin x}{x} = 1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \frac{x^6}{7!} + \ldots,$$

the last series being convergent for any $x$. Integrating it term by term from 0 to $a$, we get

$$\int_0^a \frac{\sin x}{x}\,dx = a - \frac{a^3}{3!\,3} + \frac{a^5}{5!\,5} - \frac{a^7}{7!\,7} + \ldots .$$

Using this equality, we can calculate this integral for any $a$ with any degree of accuracy.

And, finally, power series play a significant part in the approximation methods of solving differential equations.

### 3.14.6. Series of Complex Numbers

**1. Some data on complex numbers.**

**Definition.** *A complex number $z$ is an ordered pair of real numbers $(x, y)$, i.e. $z = (x, y)$, where $x$ is the real part and $y$ is the imaginary part of the complex number.*

The complex number $z = (x, y)$ is depicted on the $xy$-plane by a point $M$ with coordinates $(x, y)$ or by a vector $\overrightarrow{OM}$ whose projections on the $x$-axis and $y$-axis are $x$ and $y$ respectively (Fig. 216). In this case the $xy$-plane is called a *conditionally complex plane.*

A complex number of the form $(x, 0)$ is identified with the real number $x$, i.e. $(x, 0) = x$. This allows us to regard the set of all real numbers as a subset of the set of complex numbers. On a complex plane, real numbers are depicted by points on the $x$-axis which is known as the *real axis*.

For $y \neq 0$, the number $(x, y)$ is an *imaginary complex number*. The imaginary number $(0, y)$ is *pure imaginary* and is symbolically designated as $z = iy$. The pure imaginary number $(0, 1)$ is known as an *imaginary unit* and is designated as $i$, i.e. $i = (0, 1)$. Pure imaginary numbers are designated on the complex plane by points on the $y$-axis which is known as the *imaginary axis*.

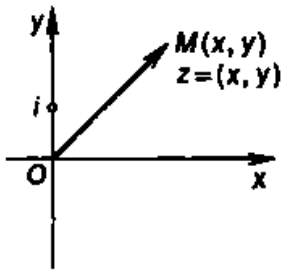


Fig. 216

Two complex numbers $z_1 = (x_1, y_1)$ and $z_2 = (x_2, y_2)$ are *equal* if $x_1 = x_2$ and $y_1 = y_2$. Hence it follows that the vectors which represent equal complex numbers are equal. The complex number $0 = (0, 0)$ is *zero*.

**Operations involving complex numbers.** Here are the definitions of algebraic operations involving complex numbers.

The *sum* of the complex numbers $z_1 = (x_1, y_1)$ and $z_2 = (x_2, y_2)$ is a complex number

$$z = (x_1 + x_2, y_1 + y_2).$$

The *product* of the complex numbers $z_1 = (x_1, y_1)$ and $z_2 = (x_2, y_2)$ is a complex number

$$z = (x_1x_2 - y_1y_2, x_1y_2 + x_2y_1).$$

The subtraction of complex numbers is defined as the inverse of addition, i.e. the number $z$ is the *difference* of the numbers $z_1$ and $z_2$ if $z + z_2 = z_1$. It follows from this definition that the complex number

$$z = (x_1 - x_2, y_1 - y_2)$$

is the difference of the complex numbers $z_1 = (x_1, y_1)$ and $z_2 = (x_2, y_2)$.

The division of complex numbers is defined as the inverse of multiplication, i.e. the number $z$ is the *quotient* of the numbers $z_1$ and $z_2 \neq 0$ if $z \cdot z_2 = z_1$. It follows from this definition that the complex number

$$z = \left(\frac{x_1x_2 + y_1y_2}{x_2^2 + y_2^2}, \frac{x_2y_1 - x_1y_2}{x_2^2 + y_2^2}\right)$$

is the quotient of the complex numbers $z_1 = (x_1, y_1)$ and $z_2 = (x_2, y_2)$.

The imaginary unit $i = (0, 1)$ plays a special part in the operations involving complex numbers. Multiplying it by itself (i.e. squaring it), we get $(0, 1) \cdot (0, 1) = (-1, 0) = -1$, i.e. $i^2 = -1$, by

virtue of the definition of the product of complex numbers. Thus any complex number $z = (x, y)$ can be represented as

$$z = (x, y) = (x, 0) + (0, y) = (x, 0) + (y, 0)\cdot(0, 1) = x + iy, \quad (1)$$

and then we can perform all the actions on complex numbers according to the ordinary rules of the algebra of polynomials. Notation (1) is the *algebraic form* of a complex number.

The complex number $\bar{z} = (x, -y) = x - iy$ is a *complex conjugate* of the number $z = (x, y) = x + iy$ and is represented on the complex plane by a point which is symmetric with respect to the point $z$ about the $x$-axis.

**Example.** Find the product of the conjugate numbers $z = x + iy$ and $\bar{z} = x - iy$.

*Solution.* We have

$$z\cdot\bar{z} = (x + iy)\cdot(x - iy) = xx + ixy - iyx + yy = x^2 + y^2.$$

**The trigonometric form of a complex number.** On the complex $xy$-plane we introduce a polar system of coordinates such that the pole is at the origin $O$ of the rectangular system and the polar axis coincides with the positive semiaxis $Ox$ (Fig. 217). We consider a complex number $z = x + iy$. From the formulas $x = \rho\cos\varphi$ and $y = \rho\sin\varphi$, which relate the polar and rectangular coordinates, we get the trigonometric form of notation of the complex number $z = x + iy$:

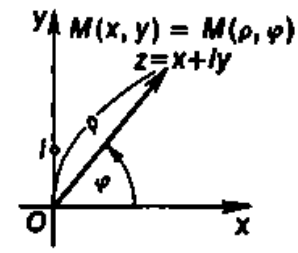


Fig. 217

$$z = \rho(\cos\varphi + i\sin\varphi).$$

The number $\rho$ is the *modulus* (the *absolute value*) and the number $\varphi$ is the *argument* of the complex number $z$. They are designated as $\rho = |z|$ and $\varphi = \operatorname{Arg} z$. Note that the argument $\varphi$ of the number $z$ is defined not uniquely but with an accuracy to within the term $2\pi n$, where $n = 0, \pm 1, \pm 2, \ldots$, and the modulus of the number $z = x + iy$ has the value

$$\rho = |z| = |\overrightarrow{OM}| = \sqrt{x^2 + y^2}.$$

The value of the argument which satisfies the inequalities $0 \leqslant \varphi < 2\pi$ is the *principal value* of the argument and is designated as $\varphi = \arg z$. The argument of the complex number $z = 0$ is not defined and its modulus is equal to zero.

**Example.** Represent the following numbers in trigonometric form: (1) $z = 2 + i2\sqrt{3}$; (2) $z = 5$; (3) $z = i$.

*Solution.* (1) For $z = 2 + i2\sqrt{3}$ we have $x = 2$, $y = 2\sqrt{3}$, $\rho = \sqrt{2^2 + (2\sqrt{3})^2} = 4$, $\cos\varphi = x/\rho = 1/2$, $\sin\varphi = y/\rho = \sqrt{3}/2$. These

values of cosine and sine are associated with the value of the argument $\varphi = \pi/3$. Consequently, $z = 2 + i2\sqrt{3} = 4[\cos(\pi/3) + i\sin(\pi/3)]$.

(2) For $z = 5$ we have $x = 5$, $y = 0$, $\rho = 5$, $\cos\varphi = 1$, $\sin\varphi = 0$. Thus $z = 5 = 5(\cos 0 + i\sin 0)$.

(3) For $z = i$ we have $x = 0$, $y = 1$, $\rho = 1$, $\cos\varphi = 0$, $\sin\varphi = 1$. Consequently, $z = i = 1[\cos(\pi/2) + i\sin(\pi/2)]$.

It is convenient to perform the operations of multiplication and division of complex numbers in trigonometric form. Let $z_1 = \rho_1(\cos\varphi_1 + i\sin\varphi_1)$ and $z_2 = \rho_2(\cos\varphi_2 + i\sin\varphi_2)$. Then we have

$$\begin{aligned} z &= \rho(\cos\varphi + i\sin\varphi) = z_1 \cdot z_2 \\ &= \rho_1(\cos\varphi_1 + i\sin\varphi_1)\cdot\rho_2(\cos\varphi_2 + i\sin\varphi_2) \\ &= \rho_1\rho_2(\cos\varphi_1\cos\varphi_2 + i\cos\varphi_1\sin\varphi_2 \\ &\quad + i\sin\varphi_1\cos\varphi_2 - \sin\varphi_1\sin\varphi_2) \\ &= \rho_1\rho_2[\cos(\varphi_1 + \varphi_2) + i\sin(\varphi_1 + \varphi_2)]. \end{aligned}$$

Thus

$$\rho = \rho_1\rho_2, \quad \varphi = \varphi_1 + \varphi_2,$$

i.e. when complex numbers are multiplied, their moduli are multiplied and the arguments are added together. When complex numbers are divided for $\rho_2 \neq 0$, there hold relations

$$\rho = \frac{\rho_1}{\rho_2}, \quad \varphi = \varphi_1 - \varphi_2.$$

When $n$ equal complex numbers $z = \rho(\cos\varphi + i\sin\varphi)$ are multiplied, we have $z^n = \rho^n(\cos n\varphi + i\sin n\varphi)$. If we set $\rho = 1$, we get a formula

$$(\cos\varphi + i\sin\varphi)^n = \cos n\varphi + i\sin n\varphi,$$

which is known as *De Moivre's* formula.*

**2. The limit of a sequence of complex numbers.** The theory of limits considered earlier for sequences of real numbers can be generalized to the case of sequences of complex numbers, and then many definitions connected with the limiting process are retained in full.

Let us consider a sequence of complex numbers $z_1, z_2, \ldots, z_n, \ldots$, where $z_n = (x_n + iy_n)$ $(n = 1, 2, \ldots)$.

The complex number $a = x_0 + iy_0$ is the *limit of the sequence* $\{z_n\}$ if for any $\varepsilon > 0$ there is a number $N$ such that the inequality $|z_n - a| < \varepsilon$ is satisfied for $n > N$. In this case the sequence $\{z_n\}$ is said to be *converging* to the number $a$, and this is written as $\lim\limits_{n\to\infty} z_n = a$, or $z_n \to a$ as $n \to \infty$.

* De Moivre, Abraham (1667-1754), an English mathematician, born in France.

In terms of geometry, this means that for any $\varepsilon > 0$, beginning with a certain number $N$ which depends on $\varepsilon$, all the elements of the sequence $\{z_n\}$ fall in a circle of radius $\varepsilon$ with centre at the point $a$, and the points $z_n$ infinitely tend to the point $a$ with an increase in $n$. This follows from the equalities $|z_n - a| = |(x_n - x_0) + i(y_n - y_0)| = \sqrt{(x_n - x_0)^2 + (y_n - y_0)^2} = \rho(M_n, M_0)$, where $\rho(M_n, M_0)$ is the distance between the points $M_n(x_n, y_n)$ and $M_0(x_0, y_0)$, i.e. between the points which represent the numbers $z_n$ and $a$.

Every complex number $z_n$ is an ordered pair of real numbers $(x_n, y_n)$ and therefore the sequence $\{z_n\}$ is associated with two sequences $\{x_n\}$ and $\{y_n\}$ of real numbers composed of the real and the imaginary parts of the elements $z_n$ of the sequence $\{z_n\}$ respectively.

Consider the following theorem.

**Theorem 14.18.** *The necessary and sufficient condition for the convergence of the sequence $\{z_n\} = \{x_n + iy_n\}$ is the convergence of the sequences $\{x_n\}$ and $\{y_n\}$ of real numbers.*

*Proof.* **Necessity.** Let $z_n \to a = x_0 + iy_0$. Then, for any $\varepsilon > 0$, there is a number $N$ such that the inequality $|z_n - a| = \sqrt{(x_n - x_0)^2 + (y_n - y_0)^2} < \varepsilon$ is satisfied for $n > N$. Hence $|x_n - x_0| \leqslant |z_n - a| < \varepsilon$ and $|y_n - y_0| \leqslant |z_n - a| < \varepsilon$ for $n > N$ and, consequently, $x_n \to x_0$ and $y_n \to y_0$.

**Sufficiency.** Assume that $x_n \to x_0$ and $y_n \to y_0$ and let $a = x_0 + iy_0$. Then, from the relation $|z_n - a| = \sqrt{(x_n - x_0)^2 + (y_n - y_0)^2}$ it follows that $z_n \to a$ as $n \to \infty$. ■

This theorem makes it possible to extend the main results of the theory of limits for sequences of real numbers to sequences of complex numbers.

**3. Number series with complex terms.** We compose a series of complex numbers

$$z_1 + z_2 + \ldots + z_n + \ldots = \sum_{n=1}^{\infty} z_n, \tag{2}$$

where $z_n = x_n + iy_n$. Series (2) is *convergent* if the sequence $\{S_n\}$ of its partial sums $S_n = z_1 + z_2 + \ldots + z_n = \left(\sum_{k=1}^{n} x_k\right) + i\left(\sum_{k=1}^{n} y_k\right)$ converges as $n \to \infty$; the limit $S = \lim_{n\to\infty} S_n$ is the *sum* of the series.

Series (2) with complex terms is associated with two real series $\sum_{n=1}^{\infty} x_n$ and $\sum_{n=1}^{\infty} y_n$. As for a sequence of complex numbers (see Theorem 14.18), we can prove that for series (2) to be convergent, it is necessary and sufficient that the series $\sum_{n=1}^{\infty} x_n$ and $\sum_{n=1}^{\infty} y_n$ be

convergent, and if $\sum_{n=1}^{\infty} x_n = S'$ and $\sum_{n=1}^{\infty} y_n = S''$, then $S = \sum_{n=1}^{\infty} z_n = S' + iS''$.

The following theorem is useful for the test for convergence of series (2).

**Theorem 14.19.** *Assume that we are given a series with complex terms*

$$z_1 + z_2 + \ldots + z_n + \ldots = \sum_{n=1}^{\infty} z_n. \tag{3}$$

*Then, if the series composed of the moduli of its terms*

$$|z_1| + |z_2| + \ldots + |z_n| + \ldots = \sum_{n=1}^{\infty} |z_n| \tag{4}$$

*is convergent, then series* (3) *is convergent too.*

*Proof.* Let $z_n = x_n + iy_n$. Then $|z_n| = \sqrt{x_n^2 + y_n^2}$ and $|x_n| \leqslant \sqrt{x_n^2 + y_n^2}$, $|y_n| \leqslant \sqrt{x_n^2 + y_n^2}$. Therefore, according to the comparison test for series of real terms (see Theorem 14.6), the convergence of series (4) implies the convergence of the series $\sum_{n=1}^{\infty} |x_n|$ and $\sum_{n=1}^{\infty} |y_n|$. Consequently, the series $\sum_{n=1}^{\infty} x_n$ and $\sum_{n=1}^{\infty} y_n$ are convergent, the convergence being absolute, and this is sufficient for the convergence of series (3). ■

If series (4) is convergent, then we say that series (3) is absolutely convergent.

The theorem we have proved allows us to employ all sufficient tests for the convergence of series with nonnegative real terms when investigating the convergence of series with complex terms.

**Example.** Test the series $\sum_{n=1}^{\infty} \frac{(1+i)^n}{n!}$ for convergence.

*Solution.* This series is absolutely convergent since, when we apply d'Alembert's test to the series composed of the moduli of its terms, we obtain

$$\lim_{n\to\infty} \frac{|z_{n+1}|}{|z_n|} = \lim_{n\to\infty} \frac{|1+i|^{n+1} \cdot n!}{(n+1)!\,|1+i|^n} = \lim_{n\to\infty} \frac{|1+i|}{n+1}$$

$$= \lim_{n\to\infty} \frac{\sqrt{1^2+1^2}}{n+1} = \lim_{n\to\infty} \frac{\sqrt{2}}{n+1} = 0 < 1.$$

**4. Power series with complex terms.** *A series of the form*

$$a_0 + a_1(z-a) + a_2(z-a)^2 + \ldots + a_n(z-a)^n + \ldots = \sum_{n=0}^{\infty} a_n(z-a)^n, \tag{5}$$

*where $z$ is a complex variable and $a_n$ and $a$ are complex numbers, is a power series.*

As for real numbers, we shall restrict our discussion to power series of the form

$$a_0 + a_1 z + a_2 z^2 + \ldots + a_n z^n + \ldots = \sum_{n=0}^{\infty} a_n z^n. \qquad (6)$$

Series (5) can be reduced to series (6) by means of the substitution $z - a = t$.

To find the domain of convergence of the power series (6), we make use of Abel's theorem (see Theorem 14.11), which is formulated and proved in the same way as for real numbers. Here is the formulation of the theorem.

**Theorem 14.20.** (1) *If the power series* (6) *converges for* $z = z_0$ $(z_0 \neq 0)$, *then it converges for all $z$ satisfying the condition* $|z| < |z_0|$, *the convergence being absolute*; (2) *if series* (6) *diverges for* $z = z_1$, *then it diverges for all $z$ satisfying the condition* $|z| > |z_1|$.

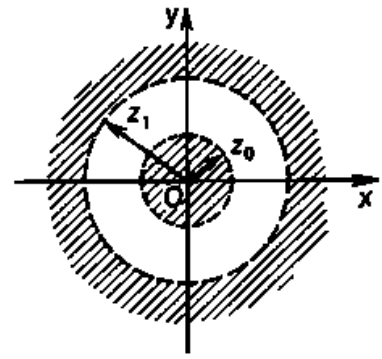


Fig. 218

Let us consider the geometric interpretation of Abel's theorem. If $z_0$ is a point of convergence of series (6), then at all points lying inside the circle of radius $|z_0|$ with centre at the origin the series is absolutely convergent, and if $z_1$ is a point of divergence of series (6), then at all points lying outside the circle of radius $|z_1|$ with centre at the origin the series is divergent (Fig. 218).

In the same way as we did for power series in the case of real numbers, we can find the shape of the domain of convergence of series (6).

**Theorem 14.21.** *If series* (6) *converges not for all values of $z$ and not only for $z = 0$, then there is a number $R > 0$ such that the series is absolutely convergent for $|z| < R$ and divergent for $|z| > R$.*

The proof of the theorem is similar to the proof of Theorem 14.12. A circle on the complex plane with centre at the origin and radius $R$ is the *circle of convergence* of the power series (6), and the number $R$ is the *radius of convergence*. At the boundary of the circle of convergence ($|z| = R$) the question about the convergence of the series is answered, as it was done in the case of real numbers, by means of an additional investigation.

If the power series (6) converges only at the point $z = 0$, then we set $R = 0$, and if the series converges for any value of $z$, i.e. throughout the complex plane, then we assume that $R = \infty$.

We can seek the radius of convergence of the power series (6) in the same way as in the case of real numbers.

**Example.** Find the radius of convergence of the power series $\sum_{n=1}^{\infty} \frac{z^n}{n!}$.

*Solution.* We have

$$R = \lim_{n\to\infty} \frac{|a_n|}{|a_{n+1}|} = \lim_{n\to\infty} \frac{(n+1)!}{n!} = \lim_{n\to\infty} (n+1) = \infty.$$

Consequently, the series is convergent for any value of $z$.

Consider a series

$$f(z) = a_0 + a_1 z + a_2 z^2 + \ldots + a_n z^n + \ldots = \sum_{n=0}^{\infty} a_n z^n. \tag{7}$$

Evidently, inside the circle of convergence of series (7) its sum $f(z)$ is a function of the complex variable $z$. Functions of a complex variable, which can be represented as power series, are *analytic functions*. We do not study functions of this kind here.*

**5. Euler's formulas.** Using power series, we shall set up two formulas which have a variety of applications.

In 3.14.5 we obtained the expansion of the function $e^x$ into a power series convergent for any value of $x$:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^n}{n!} + \ldots . \tag{8}$$

If we replace the real variable $x$ by the complex variable $z$, then we get a series in the powers of $z$:

$$1 + \frac{z}{1!} + \frac{z^2}{2!} + \ldots + \frac{z^n}{n!} + \ldots,$$

which is convergent, as follows from the example just considered, for all values of $z$. We designate its sum as $e^z$. Thus, by definition, for any complex number $z$ we have

$$e^z = 1 + \frac{z}{1!} + \frac{z^2}{2!} + \ldots + \frac{z^n}{n!} + \ldots . \tag{9}$$

The sum of series (9) is an *exponential function* of the complex variable $z$.

We can similarly define the trigonometric functions $\sin z$ and $\cos z$ of the complex variable $z$:

$$\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \ldots + (-1)^{n-1} \frac{z^{2n-1}}{(2n-1)!} + \ldots, \tag{10}$$

$$\cos z = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \ldots + (-1)^n \frac{z^{2n}}{(2n)!} + \ldots . \tag{11}$$

* See A.G. Sveshnikov and A.N. Tikhonov, *The Theory of Functions of a Complex Variable*, Mir Publishers, Moscow, 1982.

There is a simple relationship between the exponential function $e^z$ and the trigonometric functions $\sin z$ and $\cos z$. We substitute $iz$ for $z$ into (9) and on the right-hand side group all terms which contain the factor $i$ and do not contain it:

$$e^{iz} = 1 + iz + \frac{i^2z^2}{2!} + \frac{i^3z^3}{3!} + \frac{i^4z^4}{4!} + \frac{i^5z^5}{5!} + \dots$$
$$= 1 + iz - \frac{z^2}{2!} - i\frac{z^3}{3!} + \frac{z^4}{4!} + i\frac{z^5}{5!} - \dots$$
$$= \left(1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \dots\right) + i\left(z - \frac{z^3}{3!} + \frac{z^5}{5!} - \dots\right).$$

Comparing the result obtained with formulas (10) and (11), we get

$$e^{iz} = \cos z + i \sin z. \tag{12}$$

Next, substituting $-iz$ for $z$ into (9), we get

$$e^{-iz} = \cos z - i \sin z. \tag{13}$$

Formulas (12) and (13) are *Euler's formulas*. They establish a relationship between the exponential function and the trigonometric functions of the complex variable $z$. If we add and subtract relations (12) and (13) term by term, then we get another notation for Euler's formulas:

$$\cos z = \frac{e^{iz} + e^{-iz}}{2}, \quad \sin z = \frac{e^{iz} - e^{-iz}}{2i}. \tag{14}$$

Thus the functions $e^z$, $\sin z$, $\cos z$ are connected by relations (12)-(14). For $z = x$ ($x$ is a real variable), these functions of a complex variable coincide with the respective functions $e^x$, $\sin x$, $\cos x$ of a real variable.

### 3.14.7. Fourier Series

**1. The trigonometric series and its principal properties.** *A series of the form*

$$\frac{a_0}{2} + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \dots + a_n \cos nx$$
$$+ b_n \sin nx + \dots = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) \tag{1}$$

*is a trigonometric series and the numbers* $a_0$, $a_1$, $b_1$, $a_2$, $b_2$, ..., $a_n$, $b_n$, ... *are the coefficients (constants) of the trigonometric series.*

As distinct from a power series discussed earlier, in a trigonometric series we take well-known trigonometric functions

$$\frac{1}{2}, \cos x, \sin x, \cos 2x, \sin 2x, \dots, \cos nx, \sin nx, \dots \tag{2}$$

instead of the simplest functions $1, x, x^2, \dots, x^n, \dots$.

Note first of all that all functions of system (2) are *periodic* with period $2\pi$. Indeed, the constant 1/2 has any period, and the period of the functions $\sin nx$ and $\cos nx$ ($n = 1, 2, \ldots$) is $2\pi/n$* and, consequently, the number $2\pi = n\,(2\pi/n)$ is also their period. Each term of the trigonometric series (1) is evidently a periodic function with period $2\pi$, and therefore any partial sum of series (1) is $2\pi$-periodic (if all terms of the series do not change as a result of the substitution of $x + 2\pi$ for $x$, then its sum does not change as a result of this substitution either). Hence it follows that if series (1) is convergent on the interval $[-\pi, \pi]$, then it is convergent throughout the number line, and its sum, being the limit of the sequence of the periodic partial sums, is a periodic function with period $2\pi$. Trigonometric series are therefore most convenient for the study of periodic functions which describe various periodic processes occurring in nature and engineering. Oscillatory motions and rotations of various machine parts and instruments, the periodic motion of celestial bodies and elementary particles, the acoustic and electromagnetic oscillations are all examples of periodic processes.

Another important property of functions of system (2) is their *orthogonality* on the interval $[-\pi, \pi]$ in the following sense: the integral taken over the interval $[-\pi, \pi]$ of the product of any two different functions of the system is zero and the integral taken over the interval $[-\pi, \pi]$ of the square of any function of the system is nonzero.

Indeed,

$$\int_{-\pi}^{\pi} \frac{1}{2}\cos kx\,dx = \frac{1}{2k}\sin kx\Big|_{-\pi}^{\pi} = 0,$$
$$\int_{-\pi}^{\pi} \frac{1}{2}\sin kx\,dx = -\frac{1}{2k}\cos kx\Big|_{-\pi}^{\pi} = 0. \tag{3}$$

Furthermore,

$$\int_{-\pi}^{\pi} \cos kx \cos nx\,dx = \frac{1}{2}\int_{-\pi}^{\pi} [\cos(k+n)x + \cos(k-n)x]\,dx$$
$$= \frac{1}{2}\left[\frac{\sin(k+n)x}{k+n} + \frac{\sin(k-n)x}{k-n}\right]_{-\pi}^{\pi} = 0 \quad \text{for} \quad k \neq n. \tag{4}$$

By analogy we find that

$$\int_{-\pi}^{\pi} \sin kx \sin nx\,dx = 0 \ \text{ for } k \neq n, \qquad \int_{-\pi}^{\pi} \sin kx \cos nx\,dx = 0. \tag{5}$$

* Indeed, $\sin[n(x \pm 2\pi/n)] = \sin(nx \pm 2\pi) = \sin nx$.

And, finally, we have

$$\left.\begin{aligned}
&\int_{-\pi}^{\pi} \cos^2 kx\,dx = \frac{1}{2}\int_{-\pi}^{\pi}(1+\cos 2kx)\,dx\\
&= \frac{1}{2}\left[x+\frac{1}{2k}\sin 2kx\right]_{-\pi}^{\pi} = \pi,\\
&\int_{-\pi}^{\pi} \sin^2 kx\,dx = \frac{1}{2}\int_{-\pi}^{\pi}(1-\cos 2kx)\,dx\\
&= \frac{1}{2}\left[x-\frac{1}{2k}\sin 2kx\right]_{-\pi}^{\pi} = \pi,\\
&\int_{-\pi}^{\pi}\left(\frac{1}{2}\right)^2 = \frac{1}{4}x\Big|_{-\pi}^{\pi} = \frac{\pi}{2},
\end{aligned}\right\} \tag{6}$$

and this is what we wished to prove.

**2. Fourier* series.** By analogy with a power series, the following theorem is valid for a trigonometric series.

**Theorem 14.22.** *If the function $f(x)$, which is defined and integrable on the interval $[-\pi, \pi]$, can be expanded into a trigonometric series*

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx), \tag{7}$$

*which can be integrated term by term, then the expansion is unique.*

*Proof.* Integrating (7), we obtain

$$\int_{-\pi}^{\pi} f(x)\,dx = \frac{a_0}{2}\int_{-\pi}^{\pi} dx + \sum_{n=1}^{\infty}\left(a_n\int_{-\pi}^{\pi}\cos nx\,dx + b_n\int_{-\pi}^{\pi}\sin nx\,dx\right),$$

whence, taking (3) into account, we get

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,dx. \tag{8}$$

To find the coefficient $a_k$ in $\cos kx$ ($k$ is a natural number), we multiply equality (7) by $\cos kx$ and integrate with respect to $x$ from

* Fourier, Jean Baptiste Joseph (1768-1830), a French mathematician and physicist.

$-\pi$ to $\pi$*. Then, on the basis of formulas (3)-(6) we obtain

$$\int_{-\pi}^{\pi} f(x)\cos kx\,dx = \frac{a_0}{2}\int_{-\pi}^{\pi}\cos kx\,dx$$

$$+\sum_{n=1}^{\infty}\left(a_n\int_{-\pi}^{\pi}\cos kx\cos nx\,dx + b_n\int_{-\pi}^{\pi}\cos kx\sin nx\,dx\right)$$

$$= a_k\int_{-\pi}^{\pi}\cos^2 kx\,dx = a_k\pi,$$

whence it follows that

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos kx\,dx. \tag{9}$$

Similarly, multiplying relation (7) by $\sin kx$ and integrating in the limits from $-\pi$ to $\pi$, we get, on the basis of the same formulas,

$$\int_{-\pi}^{\pi} f(x)\sin kx\,dx = b_k\pi,$$

whence we find that

$$b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin kx\,dx. \tag{10}$$

Thus the coefficients $a_0$, $a_k$, $b_k$ of series (7) are uniquely defined by formulas (8)-(10), and this proves the theorem. ■

This theorem allows us to make the following definition.

**Definition.** *Let $f(x)$ be a function defined and integrable on the interval $[-\pi, \pi]$. Then the numbers $a_0$, $a_n$, $b_n$ found from formulas* (8)-(10) *are Fourier coefficients, and the series*

$$\frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n\cos nx + b_n\sin nx)$$

*with these coefficients is a Fourier series of the function $f(x)$.*

**3. Convergence of a Fourier series.** We introduce the concept of a *periodic extension* of the function $f(x)$ defined on the interval $[-\pi, \pi]$.

We say that the function $F(x)$ defined on the entire number line and periodic with period $2\pi$ is a periodic extension of the function $f(x)$ if $F(x) = f(x)$ on the interval $[-\pi, \pi]$.

* It is proved in the theory of series that series (7) can be integrated term by term after it is multiplied by a bounded function.

It is evident that if a Fourier series converges to the function $f(x)$ on the interval $[-\pi, \pi]$, then it converges to its periodic extension on the entire number line.

Let us find the conditions under which the Fourier series of the function $f(x)$ converges to this function.

The following theorem provides an answer to this question.

**Theorem 14.23.** *Let the function $f(x)$ and its derivative $f'(x)$ be continuous functions on the interval $[-\pi, \pi]$ or have a finite number of points of discontinuity of the first kind on it. Then the Fourier series of the function $f(x)$ converges throughout the number line, and at every point $x \in (-\pi, \pi)$ at which $f(x)$ is continuous the sum of the series is equal to $f(x)$, and at every point $x_0$ of discontinuity of the function the sum of the series is equal to*

$$\frac{f(x_0-)+f(x_0+)}{2},$$

*where $f(x_0-) = \lim\limits_{x\to x_0-} f(x)$ and $f(x_0+) = \lim\limits_{x\to x_0+} f(x)$. At the endpoints of the interval $[-\pi, \pi]$ the sum of the series is equal to*

$$\frac{f(-\pi)+f(\pi)}{2}.$$

*At any point $x \notin [-\pi, \pi]$ the sum of the Fourier series is equal to $F(x)$ if $x$ is a point of continuity of $F(x)$ and to $\frac{F(x-)+F(x+)}{2}$ if $x$ is a point of discontinuity of $F(x)$, where $F(x)$ is a periodic extension of $f(x)$.*

**4. Fourier series for even and odd functions.** Assume that the function $f(x)$ is defined on the interval $[-\pi, \pi]$ and is even, i.e. $f(-x) = f(x)$. Then its Fourier coefficients $b_n$ are zero. Indeed,

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,dx = \frac{1}{\pi}\left[\int_{-\pi}^{0} f(x)\sin nx\,dx + \int_{0}^{\pi} f(x)\sin nx\,dx\right].$$

We make a change of variable in the first integral in brackets. If we set $x = -t$, then $dx = -dt$; if $x = -\pi$, then $t = \pi$; if $x = 0$, then $t = 0$. Bearing in mind that the function $f(x)$ is even and the function $\sin x$ is odd, we obtain

$$\int_{-\pi}^{0} f(x)\sin nx\,dx = -\int_{\pi}^{0} f(-t)\sin n(-t)\,dt = -\int_{0}^{\pi} f(t)\sin nt\,dt.$$

Consequently

$$b_n = \frac{1}{\pi}\left[-\int_{0}^{\pi} f(t)\sin nt\,dt + \int_{0}^{\pi} f(x)\sin nx\,dx\right] = 0$$

(recall that the definite integral does not depend on the designation of the integration variable).

Similarly, bearing in mind that the functions $f(x)$ and $\cos x$ are even, we can get the following expressions for the coefficients $a_0$ and $a_n$:

$$a_0=\frac{2}{\pi}\int_0^{\pi} f(x)\,\mathrm{d}x,\quad a_n=\frac{2}{\pi}\int_0^{\pi} f(x)\cos nx\,\mathrm{d}x. \tag{11}$$

Assume now that the function $f(x)$ defined on the interval $[-\pi, \pi]$ is odd, i.e. $f(x) = -f(-x)$. Then, using arguments analogous to those presented above, we can show that the Fourier coefficients $a_n$ are zero and the coefficients $b_n$ are defined by the expressions of the form

$$b_n=\frac{2}{\pi}\int_0^{\pi} f(x)\sin nx\,\mathrm{d}x. \tag{12}$$

Thus, if the function $f(x)$ is even, then the Fourier series contains only cosines, and if the function $f(x)$ is odd, then the Fourier series contains only sines. Formulas (11) and (12) make it possible to simplify the computations of the Fourier coefficients when the given function is even or odd.

**Example 1.** Consider a function $f(x) = x$. It satisfies the conditions of Theorem 14.23 and, consequently, can be expanded into a Fourier series. Since it is odd, its Fourier coefficients $a_n = 0$ and $b_n$ can be found from formula (12). We have

$$b_n=\frac{2}{\pi}\int_0^{\pi} x\sin nx\,\mathrm{d}x$$

$$=\frac{2}{\pi}\left(-\frac{1}{n}x\cos nx\Big|_0^{\pi}+\frac{1}{n}\int_0^{\pi}\cos nx\,\mathrm{d}x\right)=(-1)^{n+1}\frac{2}{n}.$$

Thus we obtain the Fourier series of this function

$$x=2\left[\frac{\sin x}{1}-\frac{\sin 2x}{2}+\frac{\sin 3x}{3}-\frac{\sin 4x}{4}+\ldots+(-1)^{n+1}\frac{\sin nx}{n}+\ldots\right].$$

This equality is valid for any $x \in (-\pi, \pi)$. According to Theorem 14.23, at the points $x = \pm\pi$ the sum of the Fourier series does not coincide with the values of the function $f(x) = x$ and is equal to $\frac{f(-\pi)+f(\pi)}{2} = \frac{-\pi+\pi}{2} = 0$. Outside the interval $[-\pi, \pi]$ the

sum of the series is a periodic extension of the function $f(x) = x$. The graph of the sum is given in Fig. 219*a*.

**Example 2.** Consider a function $f(x) = x^2$. It satisfies the conditions of Theorem 14.23 and, consequently, can be expanded into a

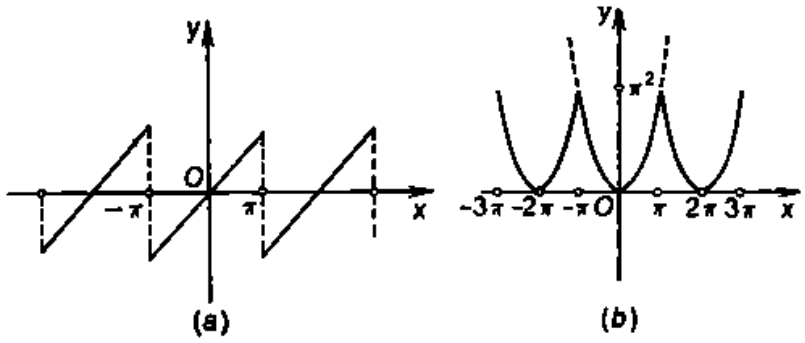



Fig. 219

Fourier series. Since it is even, its Fourier coefficients $b_n = 0$, and $a_0$ and $a_n$ can be found from formulas (11). We have

$$a_0 = \frac{2}{\pi}\int_0^\pi x^2\,dx = \frac{2\pi^2}{3}, \quad a_n = \frac{2}{\pi}\int_0^\pi x^2\cos nx\,dx$$

$$= \frac{2}{\pi}\left(\frac{x^2\sin nx}{n}\Big|_0^\pi - \frac{2}{n}\int_0^\pi x\sin nx\,dx\right) = (-1)^n\frac{4}{n^2}.$$

Hence the Fourier series of this function has the form

$$x^2 = \frac{\pi^2}{3} - 4\left(\frac{\cos x}{1} - \frac{\cos 2x}{2^2} + \frac{\cos 3x}{3^2} - \dots\right).$$

This equality is valid for any $x \in [-\pi, \pi]$ since in this case, at the points $x = \pm\pi$, the sum of the series coincides with the values of the function $f(x) = x^2$ because $[f(-\pi) + f(\pi)]/2 = (\pi^2 + \pi^2)/2 = \pi^2 = f(\pi) = f(-\pi)$. The graphs of the function $f(x) = x^2$ (shown by dotted lines) and of the sum (shown by solid lines) of this Fourier series are given in Fig. 219*b*.

**5. Fourier series with period $2l$.** Let the function $f(x)$ be defined on the interval $[-l, l]$ ($l$ is an arbitrary positive number) and satisfy the conditions of Theorem 14.23 on this interval. We expand it into a Fourier series.

We introduce a new variable $\xi$ using the formula

$$x = \frac{l\xi}{\pi}$$

and consider a function $\varphi(\xi) = f\left(\frac{l\xi}{\pi}\right) = f(x)$.

The function $\varphi(\xi)$ is evidently defined on the interval $[-\pi, \pi]$ and satisfies the conditions of Theorem 14.23 on it.

On the interval $[-\pi, \pi]$ we expand the function $\varphi(\xi)$ into a Fourier series

$$\varphi(\xi) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos n\xi + b_n \sin n\xi), \tag{13}$$

where

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} \varphi(\xi)\, d\xi, \quad a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} \varphi(\xi) \cos n\xi\, d\xi, \quad b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} \varphi(\xi) \sin n\xi\, d\xi.$$

Now we return to the old variable $x$: $x = \frac{l}{\pi}\xi$, $\xi = x\frac{\pi}{l}$, $d\xi = \frac{\pi}{l}dx$. Then formula (13) assumes the form

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{l} + b_n \sin \frac{n\pi x}{l}\right), \tag{14}$$

where

$$a_0 = \frac{1}{l} \int_{-l}^{l} f(x)\, dx, \quad a_n = \frac{1}{l} \int_{-l}^{l} f(x) \cos \frac{n\pi x}{l}\, dx,$$

$$b_n = \frac{1}{l} \int_{-l}^{l} f(x) \sin \frac{n\pi x}{l}\, dx.$$

Formula (14) is exactly the Fourier series with period $2l$.

**Example 3.** Expand the function $f(x)$ which is defined by the formula $f(x) = |x|$ on the interval $[-l, l]$ into a Fourier series with period $2l$.

*Solution.* Since the function $f(x) = |x|$ is even, we have

$$b_n = 0, \quad a_0 = \frac{2}{l} \int_0^l x\, dx = \frac{2}{l} \cdot \frac{x^2}{2}\Big|_0^l = l, \quad a_n = \frac{2}{l} \int_0^l x \cos \frac{n\pi x}{l}\, dx$$

$$= \frac{2}{l}\left(\frac{l \cdot x \sin \frac{n\pi x}{l}}{n\pi}\Bigg|_0^l - \frac{l}{n\pi} \int_0^l \sin \frac{n\pi x}{l}\, dx\right) = \frac{2l}{n^2\pi^2} \cos \frac{n\pi x}{l}\Big|_0^l$$

$$= \frac{2l}{n^2\pi^2}[(-1)^n - 1] = \begin{cases} 0 & \text{for an even } n. \\ -\frac{4l}{n^2\pi^2} & \text{for an odd } n. \end{cases}$$

Consequently, the Fourier series of the function $f(x)$ has the form

$$|x| = \frac{l}{2} - \frac{4l}{\pi^2}\left(\cos \frac{\pi x}{l} + \frac{1}{3^2} \cos \frac{3\pi x}{l} + \frac{1}{5^2} \cos \frac{5\pi x}{l} + \ldots\right).$$

The function $|x|$ satisfies the conditions of Theorem 14.23 and the relation obtained is valid for any $x \in [-l, l]$, and this means that

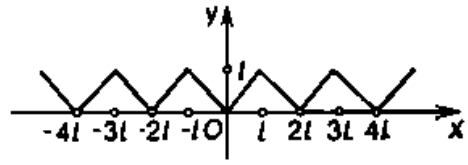



Fig. 220

the series converges throughout the number line and its sum is the function whose graph is shown in Fig. 220.

Note that Fourier series have very many applications both in theoretical research and in practical problems.

**Chapter 15**

## ORDINARY DIFFERENTIAL EQUATIONS

Differential equations occupy a place of their own in mathematics. Mathematical research of various natural phenomena often leads to the solution of these equations since the laws which a particular phenomenon obeys are themselves written as differential equations.

Differential equations are equations in which the unknown function appears under the sign of a derivative. The main task of the theory of differential equations is to study functions which are solutions of this kind of equations.

Differential equations can be classified as ordinary differential equations, in which the unknown functions are functions of one variable, and partial differential equations, in which the unknown functions are functions of two or more variables.

The theory of partial differential equations is more complicated and is discussed in more complete and special courses in mathematics. In this chapter we present elements of the theory of ordinary differential equations. In what follows, when we speak of differential equations, we mean only ordinary differential equations.

### 3.15.1. First-Order Differential Equations

We begin studying the theory of differential equations with first-order differential equations which are the simplest equations of this kind.

**1. Definition of the first-order differential equation.**

**Definition 1.** *An equation of the form*

$$F(x, y, y') = 0, \tag{1}$$

*where $x$ is an independent variable, $y$ is the sought-for function, and $y'$ is its derivative, is a first-order differential equation.*

If equation (1) can be solved for $y'$, then it assumes the form

$$y' = f(x, y) \tag{2}$$

and is a *first-order equation resolved for the derivative.* We shall exclusively deal with equations of this kind.

In some cases it is convenient to write equation (2) in the form $\frac{dy}{dx} = f(x, y)$ or in the form $f(x, y)\,dx - dy = 0$, which is a special case of the more general equation

$$P(x, y)\,dx + Q(x, y)\,dy = 0, \tag{3}$$

where $P(x, y)$ and $Q(x, y)$ are known functions. The symmetric form (3) of the equation is convenient in that the variables $x$ and $y$ are equivalent, i.e. each of them can be considered to be a function of the other.

Here are some examples of differential equations of forms (2) and (3):

$$y' = xe^y, \quad y' = \frac{y \ln x}{x}, \quad y' = x + y, \quad x\,dx + y\,dy = 0.$$

**2. Solution of an equation. Cauchy's problem.**

**Definition 2.** *The solution of a first-order differential equation is a function $y = \varphi(x)$, $x \in (a, b)$*, which, being substituted into the equation, turns it into an identity.*

Thus, for instance, the function $y = x^3$, $x \in (-\infty, +\infty)$ is a solution of the equation $3y - xy' = 0$, i.e. being substituted into the equation, it turns it into an identity, i.e. $3x^3 - x3x^2 \equiv 0$.

The graph of the solution of a differential equation is called an *integral curve.*

The answer to the question as to the conditions under which equation (2) has a solution is supplied by Cauchy's theorem which is known as the theorem on the existence and uniqueness of the solution of the differential equation (2) and is the fundamental theorem in the theory of differential equations.

**Theorem 15.1 (Cauchy's theorem).** *If the function $f(x, y)$ and its partial derivative $f'_y(x, y)$ are defined and continuous in a domain $G$ of the $xy$-plane, then, whatever the interior point $(x_0, y_0)$ of the domain $G$, there is a unique solution of the equation $y' = f(x, y)$ in some neighbourhood of this point, and this solution satisfies the conditions*

$$y = y_0 \text{ for } x = x_0. \tag{4}$$

* The interval may be finite or infinite to one side or to both sides.

Cauchy's theorem allows us to decide whether a solution exists and is unique proceeding from the form of the differential equation (2). This is especially important in the cases when it is not known beforehand whether the equation has a solution.

In terms of geometry, the theorem states that a unique integral curve passes through every interior point $(x_0, y_0)$ of the domain $G$. Equation (2) evidently has an infinite number of various solutions in the domain $G$.

Conditions (4), by virtue of which the function $y = \varphi(x)$ assumes the given value $y_0$ at the given point $x_0$, are the *initial conditions* of the solution and are usually written as

$$y|_{x=x_0} = y_0. \tag{5}$$

Finding the solution of equation (2), which satisfies the initial conditions (5), is one of the most important problems of the theory of differential equations. It is known as *Cauchy's problem*. From the point of view of geometry, to solve Cauchy's problem is to isolate, from the multitude of integral curves, the curve which passes through the given point $(x_0, y_0)$ of the $xy$-plane.

The points of the plane through which either more than one integral curve pass or not a single integral curve passes are known as *singular points* of that equation.

**3. The general and a particular solution of an equation.** We give two main definitions.

**Definition 3.** *The general solution of equation* (2) *in a domain $G$ of the $xy$-plane is a function $y = \varphi(x, C)$ which depends on $x$ and on the arbitrary constant $C$ if it is a solution of equation* (2) *for any value of the constant $C$ and if for any initial conditions* (5) *such that $(x_0, y_0) \in G$ there is a unique value of the constant $C = C_0$ such that the function $y = \varphi(x, C_0)$ satisfies the given initial conditions $\varphi(x_0, C_0) = y_0$.*

**Definition 4.** *A particular solution of equation* (2) *in the domain $G$ is a function $y = \varphi(x, C_0)$ which results from the general solution $y = \varphi(x, C)$ for a definite value of the constant $C = C_0$.*

In terms of geometry, the general solution $y = \varphi(x, C)$ is a family of integral curves on the $xy$-plane which depends on one arbitrary constant $C$, and the particular solution $y = \varphi(x, C_0)$ is one integral curve of this family which passes through the given point $(x_0, y_0)$.

Sometimes the initial conditions (5) are called *Cauchy's conditions*, and the *solution of one of Cauchy's problems* is called a particular solution.

**Example 1.** Consider an equation $y' = 3x^2$.

This is a first-order differential equation. It satisfies all conditions of Cauchy's theorem since the functions $f(x, y) = 3x^2$ and $f'_y(x, y) = 0$ are defined and continuous on the entire $xy$-plane. It is easy to verify that the function $y = x^3 + C$, where $C$ is an arbitrary con-

stant, is the general solution of this equation on the entire $xy$-plane.

In terms of geometry, this general solution is a family of cubic parabolas. For different values of the constant $C$ we get different solutions of this equation. For instance, if $C = 0$, then $y = x^3$, if $C = -1$, then $y = x^3 - 1$, if $C = 2$, then $y = x^3 + 2$ and so on.

To solve a Cauchy's problem, i.e. to find a particular solution, we specify arbitrary initial conditions: $x = x_0$ and $y = y_0$. Substituting these values into the general equation $y = x^3 + C$ for $x$ and $y$, we obtain $y_0 = x_0^3 + C$, whence we find that $C = y_0 - x^3$. We

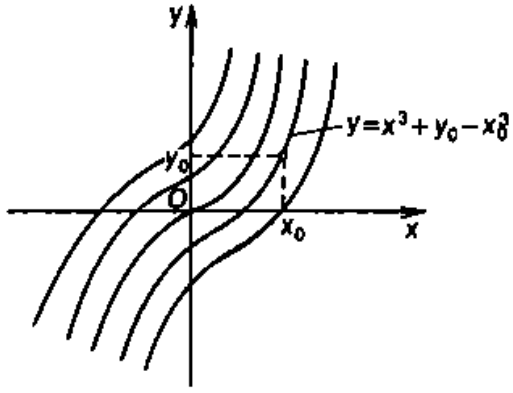



Fig. 221

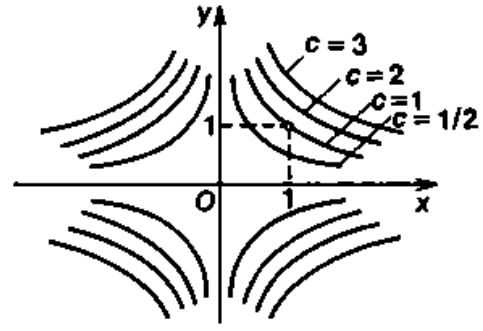



Fig. 222

have thus found a particular solution $y = x^3 + y_0 - x_0^3$. In terms of geometry, this means that from the family of cubic parabolas $y = x^3 + C$ we have chosen one parabola which passes through the given point $(x_0, y_0)$ (Fig. 221).

**Example 2.** Consider an equation $y' = -y/x$.

This is a first-order differential equation. The functions $f(x, y) = -y/x$ and $f'_y(x, y) = -1/x$ are continuous for $x \neq 0$. Consequently, this equation satisfies the conditions of Cauchy's theorem on the entire $xy$-plane, except for the $y$-axis.

It is easy to verify that the function $y = C/x$, where $C$ is an arbitrary constant, is the general solution of this equation in the domains $y > 0$ and $y < 0$. For different values of the constant $C$ we get different solutions.

Let us find a particular solution which satisfies, say, the initial conditions $x_0 = 1$ and $y_0 = 1$. We have $1 = C/1$. Hence $C = 1$, and $y = 1/x$ is the required particular solution.

In terms of geometry, the general solution of this equation is a family of hyperbolas $y = C/x$, each of which represents a particular solution of the equation. Specifying the initial conditions $x_0 = 1$ and $y_0 = 1$, we isolate, from the family, the hyperbola which passes through the point (1, 1) of the $xy$-plane (Fig. 222).

Note that not a single integral curve passes through the points of the $y$-axis, i.e. these are singular points of the given equation.

**4. Geometric meaning of an equation.** Let us be given a first-order differential equation $y' = f(x, y)$ and assume that the function $y = \varphi(x)$ is its solution. The graph of the solution is a continuous integral curve through every point of which we can draw a tangent. It follows from this equation that the slope $y'$ of the tangent to the integral curve at its every point $(x, y)$ is equal to the value of the right-hand side of the equation $f(x, y)$ at this point. Thus the equation $y' = f(x, y)$ establishes a relationship between the coordinates of the point $(x, y)$ and the slope $y'$ of the tangent to the graph of the integral curve at the same point. Knowing $x$ and $y$, we can indicate the direction of the tangent to this integral curve at the point $(x, y)$.

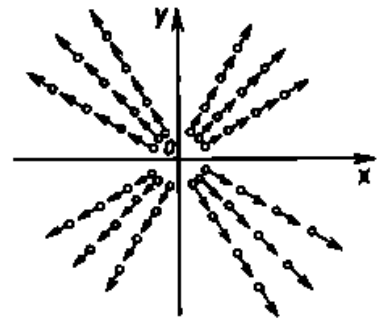


Fig. 223

Let us put every point $(x, y)$ of the integral curve into correspondence with a directed segment whose slope is $f(x, y)$. We obtain the so-called *direction field* of this equation which discloses the geometric meaning of the first-order differential equation.

Thus, from the point of view of geometry, the equation $y' = f(x, y)$ defines the direction field on the $xy$-plane, and the solution of the equation is the integral curve the direction of the tangent to which at every point is coincident with the direction of the field at that point.

Constructing the direction field of the given differential equation on the plane, we can approximate the integral curves.

**Example 3.** Consider an equation $y' = y/x$.

The function $f(x, y) = y/x$ is not defined for $x = 0$ and, consequently, the direction field of this equation can be constructed on the entire plane, except for the $y$-axis.

At every point $(x, y)$ $(x \neq 0)$ the slope $y'$ of the tangent to the integral curve is equal to $y/x$ and is coincident with the slope of the straight line passing through the origin and this point. Figure 223 shows the direction field of this equation. Evidently the straight lines $y = Cx$ ($C$ is an arbitrary constant) are the integral curves.

Let us discuss now the methods of finding the solutions of first-order differential equations. Note that there is no general method of finding solutions. We usually consider equations of different types and for each of them find a means of its solution.

**5. Equations with variables separable.**

**Definition 5.** *An equation of the form*

$$y' = f_1(x) f_2(y), \tag{6}$$

*where $f_1(x)$ and $f_2(y)$ are continuous functions, is known as a differential equation with variables separable.*

To find the solution of equation (6), we must, as they say, separate the variables in it. For that purpose, we substitute $\frac{dy}{dx}$ for $y'$ in (6), divide both sides of the equation by $f_2(y)$ (we assume that $f_2(y) \neq 0$) and multiply by $dx$. Then equation (6) assumes the form

$$\frac{dy}{f_2(y)} = f_1(x)\,dx. \tag{7}$$

In this equation the variable $x$ appears only on the right-hand side and the variable $y$ only on the left-hand side (i.e. the variables are separated).

Assuming that the function $y = \varphi(x)$ is a solution of the equation and substituting it into (7), we get an identity. Integrating the identity, we get

$$\int \frac{dy}{f_2(y)} + C_1 = \int f_1(x)\,dx + C_2,$$

or

$$\int \frac{dy}{f_2(y)} = \int f_1(x)\,dx + C,\,* \tag{8}$$

where $C = C_2 - C_1$ is an arbitrary constant.

Relation (8) implicitly defines the general solution of equation (6).

**Example 4.** Solve the equation $y' = y/x$ (compare with Example 3).

*Solution.* This equation is of form (6), where $f_1(x) = 1/x$ and $f_2(y) = y$. Separating the variables, we get $\frac{dy}{y} = \frac{dx}{x}$. Integrating, we obtain

$$\int \frac{dy}{y} = \int \frac{dx}{x} + \ln|C_1|, \quad C_1 \neq 0,^{**} \text{ or } \ln|y| = \ln|x| + \ln|C_1|.$$

Taking antilogarithms, we find that $|y| = |C_1|\,|x|$, and this is equivalent to the equation $y = \pm C_1 x$. Setting $\pm C_1 = C$, we finally obtain

$$y = Cx \tag{9}$$

which is the general solution of this equation, where $C$ is an arbitrary constant which can assume positive as well as negative values, but $C \neq 0$. Note that $y = 0$ is also a solution of the equation (we have lost it when we divided by $y$). We can include this solution into (9) if we assume that the constant $C$ also assumes the value $C = 0$.

* In the theory of differential equations the symbol of the indefinite integral denotes not the whole set of antiderivatives but one of the antiderivatives of that set.

** To simplify the notation, we have designated the arbitrary constant not as $C$ but as $\ln|C_1|$. It is possible to do this since $\ln|C_1|$ can assume any value from $-\infty$ to $+\infty$.

In terms of geometry, the general solution (9) is a family of straight lines which pass through the origin.

Assume that we have to isolate from the general solution (9) a particular solution satisfying the following initial conditions: $x_0 = 1$ and $y_0 = 2$. Substituting these values into the general solution (9) for $x$ and $y$, we get $2 = C \cdot 1$, whence $C = 2$. Thus the required particular solution is $y = 2x$.

**6. Linear equations.**

**Definition 6.** *An equation of the form*

$$y' + p(x)\, y = f(x), \tag{10}$$

*where $p(x)$ and $f(x)$ are continuous functions, is a linear differential equation of the first order.*

The name of the equation can be explained by the fact that the unknown function $y$ and its derivative $y'$ are linear in the equation, i.e. of the first order.

If $f(x) \equiv 0$, then equation (10) is a *homogeneous linear equation.* If $f(x) \not\equiv 0$, then equation (10) is a *nonhomogeneous linear equation.*

To find the general solution of equation (10), use can be made of the *method of variation of the constant.*

Using this method, we first find the general solution of the homogeneous linear equation

$$y' + p(x)\, y = 0, \tag{11}$$

which corresponds to the nonhomogeneous equation (10). Equation (11) is an equation with variables separable. Separating the variables and integrating, we obtain

$$\frac{dy}{y} = -p(x)\, dx, \quad \ln|y| = -\int p(x)\, dx + \ln|C_1|.$$

From this, taking antilogarithms, we find the general solution of equation (11):

$$y = \pm C_1 e^{-\int p(x)\, dx}, \text{ or } y = Ce^{-\int p(x)\, dx}, \tag{12}$$

where $C = \pm C_1$ is an arbitrary constant.

We seek now the general solution of equation (10) in form (12), where we assume $C$ to be not a constant but a new unknown function of $x$ (here is the sense of the method), i.e. in the form

$$y = C(x)\, e^{-\int p(x)\, dx}. \tag{13}$$

To find the function $C(x)$ and thus find the solution in form (13), we substitute function (13) into equation (10). We get

$$C'(x)\, e^{-\int p(x)\, dx} - C(x)\, p(x)\, e^{-\int p(x)\, dx} + C(x)\, p(x)\, e^{-\int p(x)\, dx} = f(x),$$

or

$$C'(x) = f(x)\,e^{\int p(x)\,dx}. \tag{14}$$

Thus, for function (13) to be a solution of equation (10), the function $C(x)$ must satisfy equation (14). Integrating it, we get

$$C(x) = \int f(x)\,e^{\int p(x)\,dx}\,dx + C_1,$$

where $C_1$ is an arbitrary constant. Substituting the expression found for $C(x)$ into relation (13), we get the general solution of the linear equation (10):

$$y(x) = C_1 e^{-\int p(x)\,dx} + e^{-\int p(x)\,dx} \int f(x)\,e^{\int p(x)\,dx}\,dx. \tag{15}$$

When solving specific problems, it is easier to repeat the calculations presented above every time rather than use the cumbersome formula (15).

**Example 5.** Find the general solution of the equation $y' + 3y = e^{2x}$.

*Solution.* This equation is linear. Here $p(x) = 3$ and $f(x) = e^{2x}$. We first solve the corresponding homogeneous equation $y' + 3y = 0$. Separating the variables $\frac{dy}{y} = -3dx$ and integrating, we find that

$$\ln|y| = -3x + \ln|C_1|, \text{ or } y = \pm C_1 e^{-3x} = Ce^{-3x}.$$

We seek the general solution of the given equation in the form $y = C(x)e^{-3x}$. Differentiating, we have $y' = C'(x)\,e^{-3x} - 3C(x)e^{-3x}$. Substituting the expressions for $y$ and $y'$ into the given equation, we obtain

$$C'(x)\,e^{-3x} = e^{2x},\ C'(x) = e^{5x}, \text{ or } dC = e^{5x}\,dx,$$

whence we have $C(x) = e^{5x}/5 + C_2$, where $C_2$ is an arbitrary constant. Consequently, the general solution of the given equation has the form

$$y = C(x)\,e^{-3x} = \left(\frac{1}{5}e^{5x} + C_2\right)e^{-3x}, \text{ or } y = \frac{1}{5}e^{2x} + C_2 e^{-3x}.$$

**7. Exact differential equations.**

**Definition 7.** *An equation of the form*

$$P(x, y)\,dx + Q(x, y)\,dy = 0, \tag{16}$$

*where the left-hand side is the total differential of some function $F(x, y)$ in a domain $G$, is known as an exact differential equation.*

If equation (16) is an exact differential equation, we can write it as follows:

$$dF(x, y) = 0,$$

where $F(x, y)$ is a function such that $dF(x, y) = P(x, y)\,dx + Q(x, y)\,dy$. Hence it follows that the general solution of equation (16) in implicit form is defined by the equation

$$F(x, y) = C,$$

where $C$ is an arbitrary constant. Indeed, if $y = \varphi(x)$ is a solution of equation (16), then $dF(x, \varphi(x)) \equiv 0$, i.e. $F(x, \varphi(x)) = C$, and, conversely, for any function $y = \varphi(x)$, which turns the equation $F(x, y) = C$ into an identity, we get $dF(x, \varphi(x)) = 0$, i.e. $y = \varphi(x)$ is a solution of equation (16).

Thus, the finding of the solution of equation (16) reduces to the finding of the function $F(x, y)$ whose differential is equal to $P(x, y)\,dx + Q(x, y)\,dy$.

We know (see Theorem 13.7) that for the expression $P(x, y)\,dx + Q(x, y)\,dy$ to be the total differential of some function $F(x, y)$, it is necessary and sufficient that

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}. \tag{17}$$

We assume that condition (17) is fulfilled. Then there is a function $F(x, y)$ such that $dF = \frac{\partial F}{\partial x}\,dx + \frac{\partial F}{\partial y}\,dy = P(x, y)\,dx + Q(x, y)\,dy$. Hence

$$\frac{\partial F}{\partial x} = P(x, y), \quad \frac{\partial F}{\partial y} = Q(x, y). \tag{18}$$

Integrating the relation $\frac{\partial F}{\partial x} = P(x, y)$ with respect to $x$, we obtain

$$F(x, y) = \int P(x, y)\,dx + C(y), \tag{19}$$

where $C(y)$ is an arbitrary function of $y$. We choose now a function $C(y)$ such that the second of relations (18) is fulfilled. For that purpose, we differentiate the right-hand side of relation (19) with respect to $y$ and equate the derivative to $Q(x, y)$:

$$\frac{\partial}{\partial y}\left[\int P(x, y)\,dx\right] + C'(y) = Q(x, y). \tag{20}$$

Then we find $C'(y)$ from equation (20) obtained and, integrating, get $C(y)$. Substituting the function $C(y)$ obtained into relation (19), we get the required function $F(x, y)$.

To isolate a particular solution which satisfies the initial conditions $x = x_0$ and $y = y_0$ from the general solution, we must replace $x$ and $y$ in the general solution $F(x, y) = C$ by the initial values. Then $C = F(x_0, y_0)$ and $F(x, y) = F(x_0, y_0)$ is the required particular solution.

**Example 6.** Find the general solution of the equation $(x + y + 1)\,dx + (x - y^2 + 3)\,dy = 0$ and isolate a particular solution which satisfies the initial conditions $x_0 = 1$ and $y_0 = 0$ from it.

*Solution.* Here $P(x, y) = x + y + 1$ and $Q(x, y) = x - y^2 + 3$. Since

$$\frac{\partial(x+y+1)}{\partial y} = 1 = \frac{\partial(x-y^2+3)}{\partial x},$$

the expression $(x + y + 1)\,dx + (x - y^2 + 3)\,dy$ is the total differential of some function $F(x, y)$. We have

$$F(x, y) = \int P(x, y)\,dx + C(y) = \int (x+y+1)\,dx + C(y)$$
$$= \frac{x^2}{2} + xy + x + C(y). \tag{21}$$

We find the function $C(y)$ using equation (20):

$$\frac{\partial}{\partial y}\left[\frac{x^2}{2} + xy + x + C(y)\right] = x - y^2 + 3,$$

$$C'(y) = -y^2 + 3, \quad \frac{dC}{dy} = -y^2 + 3, \quad dC = (-y^2 + 3)\,dy,$$

$$C(y) = \int (-y^2 + 3)\,dy + C_1 = -\frac{y^3}{3} + 3y + C_1.$$

Substituting $C(y)$ we have found into (21), we get

$$F(x, y) = \frac{x^2}{2} + xy + x - \frac{y^3}{3} + 3y + C_1.$$

The given equation assumes the form $dF(x, y) = 0$ and its general solution is defined by the equation

$$F(x, y) = C_2, \text{ or } \frac{x^2}{2} + xy + x - \frac{y^3}{3} + 3y + C_1 = C_2.$$

Setting $6(C_2 - C_1) = C_3$ ($C_3$ is an arbitrary constant), we get a final equation which implicitly defines the general solution of the initial differential equation

$$3x^2 + 6xy + 6x - 2y^3 + 18y = C_3.$$

Let us find now the value of the constant $C_3$ for which the particular solution satisfies the specified initial conditions. We have $3\cdot 1 + 6\cdot 1\cdot 0 + 6\cdot 1 - 2\cdot 0 + 18\cdot 0 = C_3$, whence $C_3 = 9$, and the required particular solution is defined by the equation

$$3x^2 + 6xy + 6x - 2y^3 + 18y = 9.$$

**8. Approximation of first-order differential equations using Euler's method.** We have considered several methods of finding exact solutions of first-order differential equations. If none of them makes it possible to achieve the aim and requires complicated computations

use is made of approximate methods of solving equations. The simplest of them is *Euler's method.*

The essence of the method is that the required integral curve, which is the graph of the particular solution, is approximated by a broken line.

Assume that we are given a differential equation

$$y' = f(x, y)$$

and initial conditions $y|_{x=x_0} = y_0$. Let us find an approximate solution of the equation of the interval $[x_0, b]$, which satisfies the specified initial conditions.

We use the points $x_0 < x_1 < x_2 < \ldots < x_n = b$ to divide the interval $[x_0, b]$ into $n$ equal parts. Let $x_1 - x_0 = x_2 - x_1 = \ldots = x_n - x_{n-1} = \Delta x$. Then we designate the approximate values of the required solution at the points $x_i$ as $y_i$ $(i=1, 2, \ldots, n)$. We draw straight lines parallel to the $y$-axis through the points of division $x_i$ (Fig. 224) and successively perform the same kind of operations.

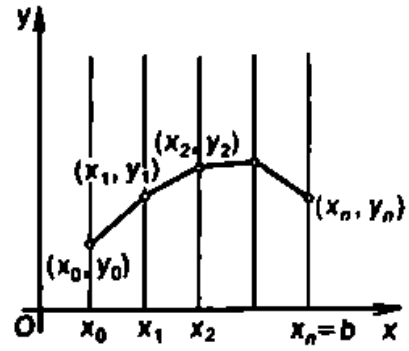



Fig. 224

We substitute the values $x_0$ and $y_0$ into the right-hand side of the equation $y' = f(x, y)$ and calculate the slope $y' = f(x_0, y_0)$ of the tangent to the integral curve at the point $(x_0, y_0)$. To find the approximate value $y_1$ of the required solution, we replace the integral curve on the interval $[x_0, x_1]$ by a segment of its tangent at the point $(x_0, y_0)$. We get

$$y_1 - y_0 = f(x_0, y_0)(x_1 - x_0),$$

whence we find, $x_0$, $x_1$, $y_0$ being known, that

$$y_1 = y_0 + f(x_0, y_0)(x_1 - x_0), \text{ or } y_1 = y_0 + f(x_0, y_0)\,\Delta x.$$

Substituting the values $x_1$ and $y_1$ into the right-hand side of the equation $y' = f(x, y)$, we calculate the slope $y' = f(x_1, y_1)$ of the tangent to the integral curve at the point $(x_1, y_1)$. Next, replacing the integral curve on the interval $[x_1, x_2]$ by a segment of the tangent, we find the approximate value $y_2$ of the solution at the point $x_2$:

$$y_2 = y_1 + f(x_1, y_1)(x_2 - x_1), \text{ or } y_2 = y_1 + f(x_1, y_1)\,\Delta x.$$

In this relation, $x_1$, $y_1$, $x_2$ are known and $y_2$ can be expressed in terms of them.

By analogy we find that

$$y_3 = y_2 + f(x_2, y_2)\,\Delta x, \ \ldots, \ y_n = y_{n-1} + f(x_{n-1}, y_{n-1})\,\Delta x.$$

We have thus approximated the required integral curve by a broken line and obtained the approximate values $y_i$ of the required solution at the points $x_i$. In this case the values $y_i$ are calculated by the formula

$$y_i = y_{i-1} + f(x_{i-1}, y_{i-1})\,\Delta x \quad (i = 1, 2, \ldots, n). \qquad (22)$$

Formula (22) is the *principal computational formula of Euler's method*. Its accuracy is the higher, the smaller the difference $\Delta x$.

In general, the degree of accuracy of Euler's method is not very high. There are more accurate methods of approximate solution of differential equations.

**Example 7.** Find the approximate solution of the equation $y' = y + x$ on the interval $[0, 1]$, which satisfies the initial conditions $x_0 = 0$ and $y_0 = 1$, and calculate $y$ for $x = 1$.

*Solution.* We divide the interval $[0, 1]$ into ten equal parts by the points $x_0 = 0$, $x_1 = 0.1$, $x_2 = 0.2$, ..., $x_{10} = 1$. We designate as $y_1, y_2, \ldots, y_{10}$ the approximate values of the solution which we shall seek using formula (22). We have

$$y_1 = 1 + (1 + 0)\cdot 0.1 = 1.1, \quad y_2 = 1.1 + (1.1 + 0.1)\cdot 0.1 = 1.22.$$

By analogy we find the other values $y_i$. It is convenient to tabulate the results of the calculations arranging them in rows:

| $x$ | $y$ | $f(x, y)\,\Delta x$ | $x$ | $y$ | $f(x, y)\,\Delta x$ |
|---|---|---|---|---|---|
| $x_0=0$ | $y_0=1$ | 0.1 | $x_6=0.6$ | $y_6=1.9431$ | 0.2543 |
| $x_1=0.1$ | $y_1=1.1$ | 0.12 | $x_7=0.7$ | $y_7=2.1974$ | 0.2897 |
| $x_2=0.2$ | $y_2=1.22$ | 0.142 | $x_8=0.8$ | $y_8=2.4871$ | 0.3287 |
| $x_3=0.3$ | $y_3=1.36$ | 0.1662 | $x_9=0.9$ | $y_9=2.8158$ | 0.3715 |
| $x_4=0.4$ | $y_4=1.5282$ | 0.1928 | $x_{10}=1$ | $y_{10}=3.1873$ | |
| $x_5=0.5$ | $y_5=1.7210$ | 0.2221 | | | |

The second column of the table includes the approximate values $y_i$ of the required solution of the given equation on $[0, 1]$ which satisfy the specified initial conditions.

The approximate value of the function $y$ for $x = 1$ is $y_{10} = 3.1873$.

To compare the approximate result to the exact one, we find the exact solution of this equation for the same initial conditions. The equation being linear, we use the method of variation of the constant. We find the general solution of the homogeneous equation

$$\frac{dy}{dx} = y, \quad \frac{dy}{y} = dx, \quad \ln|y| = x + \ln|C|, \quad y = Ce^x.$$

Varying the constant $y = C(x)\,e^x$, $y' = C'(x)\,e^x + C(x)\,e^x$ and substituting the result into the given equation, we find that

$C'(x)e^x + C(x)e^x = C(x)e^x + x$, $C'(x) = xe^{-x}$, $C(x) = -xe^{-x} - e^{-x} + C_1$,

$$y = C_1e^x - x - 1$$

is the general solution of the equation. Replacing $x$ and $y$ by the initial values $x_0 = 0$ and $y_0 = 1$, we obtain $C_1 = 2$. Thus, the exact solution of the given equation which satisfies the specified initial conditions is $y = 2e^x - x - 1$. For $x = 1$ the exact solution is $y(1) = 2(e - 1) \cong 3.4366$. Comparing it to the approximate value, we see that the absolute error is 0.2293.

**9. Some applications of first-order differential equations.** Various problems of physics involve first-order differential equations. The

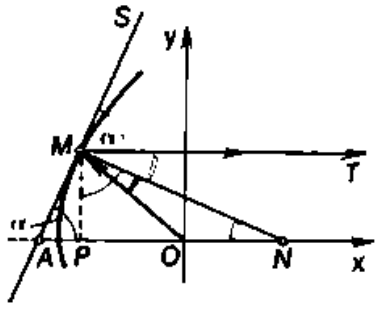



Fig. 225

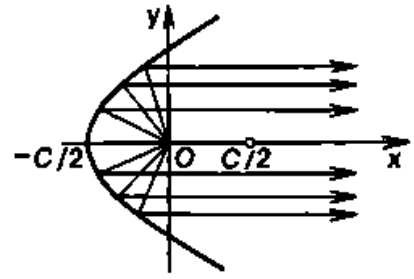



Fig. 226

main difficulty here is the derivation of differential equations since there is no universal method of doing this. Every problem requires an individual approach based on the understanding of physical laws and the skill of expressing physical problems by mathematical means.

Consider some problems of this kind.

**Problem on a searchlight.** Find the shape of the mirror which is a surface of revolution and possesses the property that all rays emanating from the source of light placed at the point $O$ on the rotation axis are reflected by the mirror parallel to that axis.

To solve the problem, we consider a plane section of the mirror passing through the rotation axis. We place the light source at the origin and make the rotation axis coincide with the $x$-axis (Fig. 225). We designate the angle the $x$-axis makes with the tangent $AS$ at the arbitrary point $M(x, y)$ of the section as $\alpha$. Our aim is to find the shape of the section, i.e. the relationship between the coordinates $y$ and $x$: $y = y(x)$. The broken line $OMT$ depicts the path of the ray from the light source at the point $O$ which is reflected from the surface of the mirror at the point $M$ parallel to the $x$-axis. We draw a normal $MN$ and drop a perpendicular $MP$ from the point $M$ to the $x$-axis. Since $\angle AMO = \angle TMS$ (the angle of incidence is equal to the reflection angle), it follows that $\angle OMN = \angle NMT = \angle ONM$. Consequently, $\triangle NOM$ is isosceles and $OM = ON$. In addition,

$\angle PMN = \alpha$ by construction. Using the relations $ON = PN - PO$, $PN = y \tan \alpha$, $PO = -x$, $ON = OM = \sqrt{x^2 + y^2}$, we arrive at a relation

$$ON = y \tan \alpha + x = \sqrt{x^2 + y^2}.$$

From this, taking into account the geometric meaning of the derivative $\left(\frac{dy}{dx} = \tan \alpha\right)$, we get the following first-order differential equation for determining the relationship between $y$ and $x$:

$$x + \frac{dy}{dx} y = \sqrt{x^2 + y^2}.$$

We transform the equation as follows: we multiply its both sides by $2dx$:

$$2x\,dx + 2y\,dy = 2\sqrt{x^2 + y^2}\,dx, \text{ or } d(x^2 + y^2) = 2\sqrt{x^2 + y^2}\,dx.$$

Using the substitution $z = x^2 + y^2$, we get an equation with variables separable

$$dz = 2\sqrt{z}\,dx,$$

which can be reduced to the form

$$z^{-1/2}\,dz = 2\,dx, \text{ whence } \sqrt{z} = x + C.$$

Replacing the variable $z$ by its expression in terms of $x$ and $y$, we obtain

$$\sqrt{x^2 + y^2} = x + C.$$

Squaring both sides of this equation, we get

$$y^2 = 2C\left(x + \frac{C}{2}\right).$$

Thus the required curve is a parabola with parameter $p = C$ and the vertex lying on the negative semi-axis $Ox$ at the distance $C/2$ from the origin (Fig. 226). Consequently, the required reflecting surface is a *paraboloid of revolution*.

**Problem on the radioactive decay.** It has been established by experiments that the rate of decay of a radioactive substance, i.e. the rate of change of its mass depending on time, is directly proportional to its amount. Let us find the law of change of the mass $m$ of the radioactive substance with time $t$, assuming that at the moment $t = 0$ the initial mass of the substance was $m_0$.

Assume that at the moment $t$ the mass of the substance is $m$, at the moment $t + \Delta t$ the mass is $m - \Delta m$. During the time $\Delta t$ the mass $\Delta m$ disintegrates. The ratio $\frac{\Delta m}{\Delta t}$ is the average rate of decay during the time $\Delta t$, and $\lim\limits_{\Delta t \to 0} \frac{\Delta m}{\Delta t} = \frac{dm}{dt}$ is the instantaneous

rate of decay at the moment $t$. According to the hypothesis,

$$\frac{dm}{dt} = -km, \tag{23}$$

where $k$ is the proportionality factor (we have taken the minus sign because the mass of the substance diminishes with time and the derivative of a decreasing function is negative). We have got a first-order differential equation from which we must find the dependence of mass $m$ on time $t$.

Solving the equation, we find that

$$\frac{dm}{m} = -k\,dt, \quad \ln m = -kt + \ln C,$$

whence

$$m = Ce^{-kt}. \tag{24}$$

Formula (24) defines the mass of a substance as a function of time. In this problem the constant $C$ has a definite value, namely, for $t = 0$ we get $m_0 = Ce^0 = C$. Substituting this value of $C$ into formula (24), we get the required dependence of the mass of the radioactive substance on time:

$$m = m_0 e^{-kt}. \tag{25}$$

Formula (24) is the general solution of the differential equation and formula (25) is a particular solution which corresponds to the initial conditions of the problem.

The coefficient $k$ can be found experimentally. For instance, for radium $k \cong 0.000447$. The time interval $T$ during which half the initial mass of the radioactive substance disintegrates is called the *half-life* of the substance. Replacing $m$ in formula (25) by the value $m_0/2$ and $k$ by the value 0.000447, we obtain an equation for finding the half-life $T$ of radium:

$$\frac{m_0}{2} = m_0 e^{-0.000447T}, \quad -0.000447T = -\ln 2,$$

whence

$$T = \frac{\ln 2}{0.000447}, \quad T \cong 1550 \text{ years}.$$

**Problem on the law of "natural growth".** The law of "natural growth" is a law according to which the rate of "growth" of a substance is directly proportional to its amount. Let us find a formula for finding the change of the amount $y$ of the substance depending on time $t$, assuming that at the initial moment $t = 0$ the amount of the substance was $y_0$.

Here $t$ is an independent variable and the amount of the substance at any time moment is the required quantity. The rate of "growth"

of the substance is the rate of change of the quantity $y$ depending on the variable $t$.

Using the physical meaning of the derivative, as we did in the preceding problem, we can write the law of "natural growth" as follows:

$$\frac{dy}{dt} = ky, \tag{26}$$

where $k > 0$ is the proportionality factor. Equation (26) differs from equation (23) only by the sign of the right-hand side and describes many processes of "multiplication".

The solution of equation (26), which satisfies the specified initial conditions $t = 0$ and $y = y_0$, has the form

$$y = y_0 e^{kt}. \tag{27}$$

Formula (27) expresses the law of "natural growth". The multiplication of neutrons in nuclear reactions, the multiplication of bacteria, the growth of crystals—all occur in accordance with this law.

### 3.15.2. Differential Equations of the Second and Higher Orders

**1. Main concepts.**

**Definition 1.** *An equation of the form*

$$F(x, y, y', y'') = 0,$$

*where $x$ is an independent variable, $y$ is the required function, $y'$ and $y''$ are its derivatives, is a second-order differential equation.*

We usually study equations which can be written in the form resolved for the second derivative:

$$y'' = f(x, y, y'). \tag{1}$$

As for a first-order differential equation, the solution of equation (1) is a *function $y = \varphi(x)$, $x \in (a, b)$, which, being substituted into the equation, turns it into an identity.* The graph of the solution is also called an *integral curve.*

The theorem on the existence and uniqueness of the solution (Cauchy's theorem), similar to the corresponding theorem for first-order equations, holds true for second-order equations.

**Theorem 15.2 (Cauchy's theorem).** *If the function $f(x, y, y')$ and its partial derivatives $f'_y(x, y, y')$ and $f'_{y'}(x, y, y')$ are defined and continuous in a domain $G$ of the space of the variables $(x, y, y')$, then, whatever the interior point $(x_0, y_0, y'_0)$ of the domain $G$, there is a unique solution of the equation $y'' = f(x, y, y')$ satisfying the conditions*

$$y = y_0, \ y' = y'_0 \ \text{for} \ x = x_0 \tag{2}$$

*in a certain neighbourhood of the point* $x_0$. In terms of geometry this means that with a point $(x_0, y_0)$ of the plane being specified, a unique integral curve, which has a definite slope $y_0'$ of the tangent at this point, passes through the point $(x_0, y_0)$.

Conditions (2) are the *initial conditions* of the solution and are often written in the form

$$y|_{x=x_0} = y_0, \quad y'|_{x=x_0} = y_0'. \tag{3}$$

As for first-order equations, the problem of finding the solution from the specified initial conditions is known as Cauchy's problem.

**2. The general and a particular solutions of an equation.** We give two main definitions.

**Definition 2.** *The general solution of equation* (1) *in a domain G of the xy-plane is a function* $y = \varphi(x, C_1, C_2)$ *which depends on x and on two arbitrary constants* $C_1$ *and* $C_2$ *if it is a solution of equation* (1) *for any values of the constants* $C_1$ *and* $C_2$ *and if, for any initial conditions* (3), *there are unique values of the constants* $C_1 = C_1^0$ *and* $C_2 = C_2^0$ *such that the function* $y = \varphi(x, C_1^0, C_2^0)$ *satisfies the given initial conditions.*

**Definition 3.** *A particular solution of equation* (1) *in the domain G is a function* $y = \varphi(x, C_1^0, C_2^0)$ *which results from the general solution* $y = \varphi(x, C_1, C_2)$ *for definite values of the constants* $C_1 = C_1^0$ *and* $C_2 = C_2^0$.

Consider an equation $y'' = 2$, for instance. This is a second-order equation. Since the functions $f(x, y, y') = 2$, $f_y'(x, y, y') = 0$, $f_{y'}'(x, y, y') = 0$ are defined and continuous in the whole space of the variables $(x, y, y')$, this equation satisfies the requirements of Cauchy's theorem in the whole space.

The general solution of this equation can be found by two-fold successive integration which first yields the first derivative $y' = 2x + C_1$ and then the general solution

$$y = x^2 + C_1x + C_2,$$

where $C_1$ and $C_2$ are arbitrary constants. In terms of geometry, the general solution is a family of parabolas, and, since it depends on two arbitrary constants, an infinite number of parabolas pass through every point of the plane and have different tangents at that point. Therefore, to isolate one parabola from the family obtained, in addition to the point $(x_0, y_0)$ through which the parabola passes, we must also define the slope $y_0'$ of the tangent to the parabola at this point.

Let us find, for instance, a particular solution of the given equation for the initial conditions $y|_{x=1} = 1$ and $y'|_{x=1} = 1$. Substituting these values into the expressions for the general solution $y = x^2 + C_1x + C_2$ and for its derivative $y' = 2x + C_1$, we

get a system of equations

$$\begin{cases} 1 = 1 + C_1 + C_2, \\ 1 = 2 + C_1 \end{cases}$$

for $C_1$ and $C_2$, whence we find that $C_1 = -1$ and $C_2 = 1$. Consequently, the required particular solution is the function

$$y = x^2 - x + 1$$

whose graph is a parabola which passes through the point (1, 1) with a slope equal to unity at this point.

**3. Second-order equations which admit a lowering of the order.** We shall consider three particular cases when by means of a change of variable the solution of equation (1) reduces to the solution of a first-order equation. Such a transformation of equation (1) is known as the *lowering of the order.*

(1) *An equation of the form* $y'' = f(x)$. The equation does not include $y$ or $y'$. We introduce a new function $z(x)$ setting $z(x) = y'$. Then $z'(x) = y''$ and the equation turns into a first-order equation $z'(x) = f(x)$ with the required function $z(x)$. Solving it, we find that $z(x) = \int f(x)\,dx + C_1$. Since $z(x) = y'$, it follows that $y' = \int f(x)\,dx + C_1$. From this, integrating once again, we get the required solution

$$y = \int \left[ \int f(x)\,dx \right] dx + C_1 x + C_2,$$

where $C_1$ and $C_2$ are arbitrary constants.

**Example 1.** Find the general solution of the equation $y'' = x$.

*Solution.* Setting $z(x) = y'$, we get a first-order equation $z'(x) = x$. Integrating it, we find that $z(x) = x^2/2 + C_1$. Replacing $z(x)$ by $y'$ and integrating once again, we find the required general solution

$$y = \int \left( \frac{x^2}{2} + C_1 \right) dx + C_2 = \frac{x^3}{6} + C_1 x + C_2.$$

(2) *An equation of the form* $y'' = f(x, y')$. The equation does not include $y$. As in the preceding case, we set $z(x) = y'$, and then $z'(x) = y''$ and the equation turns into a first-order equation with respect to $z(x)$, i.e. $z' = f(x, z)$. Solving it, we find that $z(x) = \varphi(x, C_1)$. Since $z(x) = y'$, it follows that $y' = \varphi(x, C_1)$. From this, integrating once again, we get the required solution

$$y = \int \varphi(x, C_1)\,dx + C_2,$$

where $C_1$ and $C_2$ are arbitrary constants.

**Example 2.** Find the general solution of the equation $y'' - 3y'/x = x$.

*Solution.* Setting $z(x) = y'$, we get a first-order linear equation $z' - 3z/x = x$. Solving it, we find that $z(x) = C_1x^3 - x^2$. Then $y' = C_1x^3 - x^2$ and $y = C_1x^4/4 - x^3/3 + C_2$ is the required solution.

(3) *An equation of the form* $y'' = f(y, y')$. The equation does not include $x$. We introduce a new function $z(y)$ setting $y' = z$. Then

$$y'' = \frac{d(y')}{dx} = \frac{dy'}{dy}\frac{dy}{dx} = \frac{dz}{dy}\frac{dy}{dx} = \frac{dz}{dy}z(y).$$

Substituting the expressions for $y'$ and $y''$ into this equation, we get a first-order equation with respect to $z$ as a function of $y$:

$$z\frac{dz}{dy} = f(y, z).$$

Solving it, we find that $z = \varphi(y, C_1)$. Since $z = \frac{dy}{dx}$, it follows that $\frac{dy}{dx} = \varphi(y, C_1)$. Hence $\frac{dy}{\varphi(y, C_1)} = dx$. We have got an equation with variables separable from which we find the general solution of the given equation:

$$\int \frac{dy}{\varphi(y, C_1)} = x + C_2,$$

where $C_1$ and $C_2$ are arbitrary constants.

**Example 3.** Find the general solution of the equation $yy'' - 2y'^2 = 0$.

*Solution.* Setting $y' = z(y)$ and taking into account that $y'' = z\frac{dz}{dy}$, we obtain $zy\frac{dz}{dy} - 2z^2 = 0$. This is a first-order equation with variables separable. Reducing it to the form $\frac{dz}{z} = \frac{2dy}{y}$ and integrating, we get $\ln|z| = 2\ln|y| + \ln|C_1|$, whence we have $z = C_1y^2$. Taking into account that $z = \frac{dy}{dx}$, we find that $\frac{dy}{y^2} = C_1\,dx$, whence we obtain the required solution

$$-\frac{1}{y} = C_1x + C_2, \text{ or } y = -\frac{1}{C_1x + C_2}.$$

When we cancelled $z$ out of $zy\frac{dz}{dy} - 2z^2 = 0$, we have lost the solution of the equation $z = y' = 0$, i.e. $y = C = \text{const}$. In this case it is in the general solution since it results from the general solution for $C_1 = 0$ (except for the solution $y = 0$).

**4. Differential equations of higher orders.** We shall give here only the fundamental definition and general remarks pertaining to differential equations of the $n$th order.

A differential equation of order $n$ has the form

$$F(x, y, y', \ldots, y^{(n)}) = 0$$

or, if it is resolved for the leading derivative,

$$y^{(n)} = f(x, y, y', \ldots, y^{(n-1)}). \tag{4}$$

The solution of equation (4), as of equations of the first and the second order, is the function $y = \varphi(x)$, $x \in (a, b)$, which, being substituted into the equation, turns it into an identity.

The theorem on the existence and uniqueness of the solution of equation (4) is similar to the corresponding theorems given earlier for the cases of $n = 1$ and $n = 2$.

The *general solution* of equation (4) depends on $x$ and on $n$ arbitrary constants and can be written as

$$y = \varphi(x, C_1, C_2, \ldots, C_n).$$

The solutions obtained from the general solution for definite values of the constants $C_1, C_2, \ldots, C_n$ are *particular solutions* of equations (4). To isolate a particular solution of the equation from the general solution of equation (4), we can specify the initial conditions

$$y|_{x=x_0} = y_0, \quad y'|_{x=x_0} = y_0', \quad y''|_{x=x_0} = y_0'', \quad \ldots, \quad y^{(n-1)}|_{x=x_0} = y_0^{(n-1)}. \tag{5}$$

To find the solution of equation (4) which satisfies the initial conditions (5) is to *solve Cauchy's problem* for this equation.

The simplest equation of form (4) is an equation whose right-hand side depends only on $x$, i.e. an equation of the form

$$y^{(n)} = f(x). \tag{6}$$

It is easy to solve this equation. Indeed, integrating successively $n$ times, we obtain

$$y^{(n-1)} = \int f(x)\,dx + C_1,$$

$$y^{(n-2)} = \int \left[\int f(x)\,dx + C_1\right] dx + C_2 = \int dx \int f(x)\,dx + C_1 x + C_2,$$

. . . . . .

$$y = \int dx \int dx \ldots \int f(x)\,dx + C_1 \frac{x^{n-1}}{(n-1)!} + C_2 \frac{x^{n-2}}{(n-2)!} + \ldots + C_n, \tag{7}$$

where $C_1, C_2, \ldots, C_n$ are arbitrary constants. Function (7) is the general solution of equation (6).

**Example 4.** Find the general solution of a third-order equation $y''' = e^x$ and isolate from it a particular solution which satisfies the initial conditions $y|_{x=0} = 0$, $y'|_{x=0} = 0$, $y''|_{x=0} = 1$.

*Solution.* A successive integration leads to $y'' = e^x + C_1$ and $y' = e^x + C_1 x + C_2$. Integrating once again, we get the general

solution of this equation: $y = e^x + C_1 \frac{x^2}{2} + C_2 x + C_3$, where $C_1$, $C_2$, $C_3$ are arbitrary constants.

Substituting the initial conditions into the expressions for $y$, $y'$, $y''$, we have $0 = 1 + C_3$, $0 = 1 + C_2$, $1 = 1 + C_1$, whence we find that $C_1 = 0$, $C_2 = -1$, $C_3 = -1$. Thus $y = e^x - x - 1$ is the required particular solution.

### 3.15.3. Second-Order Linear Differential Equations

In the theory of differential equations, second-order linear differential equations are very significant not only because they are of a simple and well-studied type of equations but also because many practical problems of physics, mechanics, engineering and, especially, electrical engineering involve the solution of these equations.

**1. Principal concepts.**

**Definition.** *An equation of the form*

$$y'' + p(x)\, y' + q(x)\, y = f(x), \tag{1}$$

*where $y$ is the required function and $p(x)$, $q(x)$, $f(x)$ are continuous functions on an interval $(a, b)$, is known as a second-order linear differential equation.*

If $f(x) \equiv 0$, then equation (1) is a *homogeneous linear equation.* If $f(x) \not\equiv 0$, then equation (1) is a *nonhomogeneous linear equation.*

Solving equation (1) for $y''$, i.e. $y'' = -p(x)\, y' - q(x)\, y + f(x)$, we see that it is a particular case of the equation $y'' = f(x, y, y')$ and satisfies the conditions of the theorem on the existence and uniqueness of the solution. Indeed, the function $f(x, y, y') = -p(x)\, y' - q(x)\, y + f(x)$ is continuous, being a function of three variables $x$, $y$, $y'$ (it depends on $y$ and $y'$ linearly and the functions $p(x)$, $q(x)$, $f(x)$ are continuous by the hypothesis). The partial derivatives $f'_y(x, y, y') = -q(x)$ and $f'_{y'}(x, y, y') = -p(x)$ are also continuous functions of three variables $x$, $y$, $y'$ ($p(x)$ and $q(x)$ do not depend on $y$ and $y'$ and, being functions of $x$, are continuous by the hypothesis). Therefore, for any initial conditions $y|_{x=x_0} = y_0$ and $y'|_{x=x_0} = y'_0$, where $x_0 \in (a, b)$, equation (1) has a unique solution of Cauchy's problem.

We begin studying linear differential equations with homogeneous equations.

**2. Second-order homogeneous linear differential equations.** We shall consider now certain properties of solutions of the second-order homogeneous linear equations.

**Theorem 15.3.** *If the functions $y_1(x)$ and $y_2(x)$ are solutions of the equation*

$$y'' + p(x)\, y' + q(x)\, y = 0, \tag{2}$$

*then, for any values of the constants $C_1$ and, $C_2$, the function $y = C_1y_1(x) + C_2y_2(x)$ is also a solution of equation* (2).

*Proof.* Differentiating twice the function $y = C_1y_1(x) + C_2y_2(x)$ and substituting the expressions for $y$, $y'$, $y''$ into the left-hand side of equation (2), we obtain

$$C_1y_1''(x) + C_2y_2''(x) + p(x)(C_1y_1'(x) + C_2y_2'(x)) + q(x)(C_1y_1(x) + C_2y_2(x)) = C_1[y_1''(x) + p(x)y_1'(x) + q(x)y_1(x)] + C_2[y_2''(x) + p(x)y_2'(x) + q(x)y_2(x)].$$

Since, by the hypothesis, the functions $y_1(x)$ and $y_2(x)$ are solutions of equation (2), the expressions in brackets are identically zero, and this means that the function $y = C_1y_1(x) + C_2y_2(x)$ is a solution of equation (2). ■

Thus a function of the form $y = C_1y_1(x) + C_2y_2(x)$ with arbitrary constants $C_1$ and $C_2$ is a solution of equation (2). A question naturally arises whether this solution is the general solution of equation (2). We shall prove that under certain conditions the function $y = C_1y_1(x) + C_2y_2(x)$ is the general solution of equation (2). We shall first introduce the concepts of a linear dependence and a linear independence of the functions $y_1(x)$ and $y_2(x)$.

The functions $y_1(x)$ and $y_2(x)$ are said to be *linearly dependent* on $(a, b)$ if there are numbers $\alpha_1$ and $\alpha_2$, at least one of which is nonzero, such that for any $x \in (a, b)$ there holds an equality

$$\alpha_1y_1(x) + \alpha_2y_2(x) = 0. \tag{3}$$

Evidently, if the functions $y_1(x)$ and $y_2(x)$ are linearly dependent, then they are proportional. Indeed, if $\alpha_1y_1(x) + \alpha_2y_2(x) = 0$, with $\alpha_1 \neq 0$ and $y_2(x) \neq 0$, then $\frac{y_1(x)}{y_2(x)} = -\frac{\alpha_2}{\alpha_1} = \text{const}$. The converse is also true.

The functions $y_1(x)$ and $y_2(x)$ are said to be *linearly independent* on $(a, b)$ if there are no numbers $\alpha_1$ and $\alpha_2$, at least one of which is nonzero, such that for any $x \in (a, b)$ equality (3) holds true.

In other words, equality (3) is simultaneously satisfied for all $x \in (a, b)$ only if $\alpha_1 = \alpha_2 = 0$.

Evidently, if the functions $y_1(x)$ and $y_2(x)$ are linearly independent, then their ratio $\frac{y_1(x)}{y_2(x)} \neq \text{const}$, i.e. they are not proportional. Thus, for instance, the functions $y_1(x) = x^2$ and $y_2(x) = x^3$ are linearly independent on any interval $(a, b)$ since $\frac{y_1(x)}{y_2(x)} = \frac{1}{x} \neq \text{const}$, and the functions $y_1(x) = 4x^2$ and $y_2(x) = x^2$ are linearly dependent on any interval since $\frac{y_1(x)}{y_2(x)} = 4 = \text{const}$.

Let us assume now that the functions $y_1(x)$ and $y_2(x)$ are solutions of equation (2). We can decide whether they are linearly depend-

ent or not using *Wronski's* determinant*, or *Wronskian*,

$$\begin{vmatrix} y_1 & y_2 \\ y_1' & y_2' \end{vmatrix} = y_1 y_2' - y_2 y_1', \tag{4}$$

which is a function defined on $(a, b)$ and designated as $W(y_1, y_2)$, or simply $W(x)$.

**Theorem 15.4.** *If the functions $y_1(x)$ and $y_2(x)$ are linearly dependent on $(a, b)$, then the Wronskian composed of them is zero on this interval.*

*Proof.* Since the functions $y_1(x)$ and $y_2(x)$ are linearly dependent by the hypothesis, then, by definition, there are numbers $\alpha_1$ and $\alpha_2$, at least one of which is nonzero, such that equality (3) holds true, i.e. $\alpha_1 y_1(x) + \alpha_2 y_2(x) = 0$. Let $\alpha_1 \neq 0$, for instance. Then it follows from relation (3) that

$$y_1(x) = -\frac{\alpha_2}{\alpha_1} y_2(x), \quad y_1'(x) = -\frac{\alpha_2}{\alpha_1} y_2'(x).$$

Substituting the expressions for $y_1(x)$ and $y_1'(x)$ into the Wronskian, we obtain

$$W(x) = \begin{vmatrix} y_1 & y_2 \\ y_1' & y_2' \end{vmatrix} = \begin{vmatrix} -\frac{\alpha_2}{\alpha_1} & y_2 & y_2 \\ -\frac{\alpha_2}{\alpha_1} & y_2' & y_2' \end{vmatrix} = 0. \quad \blacksquare$$

**Theorem 15.5.** *If the solutions $y_1(x)$ and $y_2(x)$ of equation (2) are linearly independent on $(a, b)$, then the Wronskian composed of them is nonzero on this interval.*

*Proof.* Let us assume the contrary, i.e. that there is a point $x_0 \in (a, b)$ at which the Wronskian $W(x_0) = 0$. We set up a system of equations

$$\begin{cases} \alpha_1 y_1(x_0) + \alpha_2 y_2(x_0) = 0, \\ \alpha_1 y_1'(x_0) + \alpha_2 y_2'(x_0) = 0 \end{cases} \tag{5}$$

in which $\alpha_1$ and $\alpha_2$ are unknown numbers. Since the determinant of this system $W(x_0) = 0$, system (5) has (see 2.10.3, case 2) a nonzero solution for $\alpha_1$ and $\alpha_2$ (i.e. at least one of them is nonzero). We consider a function

$$y(x) = \alpha_1 y_1(x) + \alpha_2 y_2(x),$$

where $\alpha_1$, $\alpha_2$ is a nonzero solution of system (5).

According to Theorem 15.3, this function is a solution of equation (2). In addition, since $\alpha_1$, $\alpha_2$ is a solution of system (5), the function $y(x)$ satisfies the zero initial conditions

$$y|_{x=x_0} = 0, \quad y'|_{x=x_0} = 0. \tag{6}$$

* Wronski, Josef Maria Hõené (1778-1853), a Polish mathematician.

These initial conditions are evidently satisfied by the solution $y(x) = 0$. By the theorem on the existence and uniqueness the solution $y(x) = 0$ is the only solution of equation (2) with the initial conditions (6). Consequently, $\alpha_1 y_1(x) + \alpha_2 y_2(x) = 0$ on the interval $(a, b)$, and this means that the functions $y_1(x)$ and $y_2(x)$ are linearly dependent on $(a, b)$, but this contradicts the hypothesis. Thus $W(x) \neq 0$ for all $x \in (a, b)$. ■

We have thus established that if the functions $y_1(x)$ and $y_2(x)$ are solutions of the homogeneous linear equation (2) on $(a, b)$, then the Wronskian composed of them is either zero on $(a, b)$ ($y_1(x)$ and $y_2(x)$ are linearly dependent) or nonzero ($y_1(x)$ and $y_2(x)$ are linearly independent).

Let us find now the conditions under which the function $y = C_1 y_1(x) + C_2 y_2(x)$ is the general solution of the homogeneous linear equation (2).

**Theorem 15.6.** *If the functions $y_1(x)$ and $y_2(x)$ are linearly independent solutions of equation (2) on $(a, b)$, then the function*

$$y = C_1 y_1(x) + C_2 y_2(x), \tag{7}$$

*where $C_1$ and $C_2$ are arbitrary constants, is the general solution of equation (2).*

*Proof.* Recall that by virtue of Theorem 15.3 the function $y = C_1 y_1(x) + C_2 y_2(x)$ is a solution of equation (2) for any values of the constants $C_1$ and $C_2$. To prove that this function is the general solution of equation (2), it is sufficient to establish the fact that a particular solution can be isolated from it, which satisfies any specified initial conditions.

Let $x_0 \in (a, b)$ and assume that

$$y|_{x=x_0} = y_0, \quad y'|_{x=x_0} = y_0' \tag{8}$$

are arbitrary initial conditions. We shall show that the constants $C_1$ and $C_2$ can be chosen such that, for these values of the constants, solution (7) is a particular solution satisfying the specified initial conditions (8).

We set up a system of equations

$$\begin{cases} y_0 = C_1 y_1(x_0) + C_2 y_2(x_0), \\ y_0' = C_1 y_1'(x_0) + C_2 y_2'(x_0) \end{cases} \tag{9}$$

in which $C_1$ and $C_2$ are unknown numbers. The determinant of this system is a Wronskian $W(x_0)$. Since the functions $y_1(x)$ and $y_2(x)$ are linearly independent on $(a, b)$ by the hypothesis, $W(x_0) \neq 0$ by virtue of Theorem 15.5. Therefore system (9) has a unique solution, which will be designated as $C_1 = C_1^0$ and $C_2 = C_2^0$. Substituting $C_1^0$ and $C_2^0$ into relation (7), we get the required particular solution of equation (2): $y = C_1^0 y_1(x) + C_2^0 y_2(x)$, which satisfies conditions (8).

And this precisely means that solution (7) is the general solution of equation (2). ■

It follows from the theorem we have proved that to find the general solution of equation (2), it is sufficient to find two linearly independent particular solutions and set up expressions (7) with arbitrary constants $C_1$ and $C_2$.

**Example 1.** Find the general solution of the equation $y'' - y = 0$.

*Proof.* We have a homogeneous linear equation. It is easy to note that its particular solutions are $y_1(x) = e^x$ and $y_2(x) = e^{-x}$. Since the Wronskian

$$W(x) = \begin{vmatrix} e^x & e^{-x} \\ e^x & -e^{-x} \end{vmatrix} = -2$$

is nonzero, the solutions are linearly independent throughout the number line. Consequently, we can write the general solution of the given equation as $y = C_1e^x + C_2e^{-x}$, where $C_1$ and $C_2$ are arbitrary constants.

We shall show in conclusion how to find the general solution of equation (2) if only one particular solution of this equation is known.

Let $y_1(x)$ be a particular solution of equation (2). We introduce a new function $z$ setting $y = y_1z$. Then $y' = y_1'z + y_1z'$ and $y'' = y_1''z + 2y_1'z' + y_1z''$. Substituting the expressions for $y$, $y'$, $y''$ into equation (2) and grouping terms, we obtain

$$[y_1'' + p(x)\, y_1' + q(x)\, y_1]\, z + z'\, [2y_1' + p(x)\, y_1] + y_1z'' = 0.$$

Since $y_1(x)$ is a solution of equation (2), the expression in the first brackets is zero and the equation assumes the form

$$z'\, [2y_1' + p(x)\, y_1] + y_1z'' = 0.$$

We can lower the order of this equation setting $z' = u(x)$, where $u(x)$ is the new required function:

$$u\, [2y_1' + p(x)\, y_1] + u'y_1 = 0.$$

We have obtained a first-order equation with respect to the function $u$ with variables separable. Solving it, we find that

$$\int \frac{du}{u} = -2\int \frac{dy_1}{y_1} - \int p(x)\,dx, \quad \ln|u| = -2\ln|y_1| - \int p(x)\,dx + \ln|C|,$$

$$u = \pm\frac{C}{y_1^2}\, e^{-\int p(x)\,dx}, \quad \text{or} \quad u = \frac{C_1}{y_1^2}\, e^{-\int p(x)\,dx},$$

where $C_1$ is an arbitrary constant. Returning to the variable $z$ and multiplying the expression for $z$ by $y_1$, we get the general solution of equation (2):

$$y = C_1y_1 \int \frac{1}{y_1^2}\, e^{-\int p(x)\,dx}\, dx + C_2y_1,$$

where $C_1$ and $C_2$ are arbitrary constants.

By way of example, find the general solution of the equation $y'' - y = 0$ yourself (see Example 1), assuming one of its particular solutions to be known.

**3. Second-order nonhomogeneous linear differential equations.** Let us discuss now the principal properties of solutions of the second-order nonhomogeneous linear differential equation (1):

$$y'' + p(x)\,y' + q(x)\,y = f(x).$$

Consider the following theorem.

**Theorem 15.7.** *The general solution of equation* (1) *is the sum of any one of its particular solutions and the general solution of the corresponding homogeneous equation.*

*Proof.* Assume that $\tilde{y}(x)$ is a particular solution of equation (1) and $Y(x) = C_1y_1(x) + C_2y_2(x)$ is the general solution of the corresponding homogeneous equation (2), where $C_1$ and $C_2$ are arbitrary constants. We shall show that the function

$$y = \tilde{y}(x) + Y(x) \tag{10}$$

is the solution of equation (1). For that purpose, we find that $y' = \tilde{y}'(x) + Y'(x)$ and $y'' = \tilde{y}''(x) + Y''(x)$ and substitute them into equation (1):

$$\tilde{y}''(x) + Y''(x) + p(x)\,[\tilde{y}'(x) + Y'(x)] + q(x)\,[\tilde{y}(x) + Y(x)]$$
$$= [Y''(x) + p(x)\,Y'(x) + q(x)\,Y(x)]$$
$$+ [\tilde{y}''(x) + p(x)\,\tilde{y}'(x) + q(x)\,\tilde{y}(x)] = 0 + f(x) = f(x).$$

It follows that the function $y = \tilde{y}(x) + Y(x)$ is indeed the solution of equation (1).

We shall show now that function (10) is the general solution of equation (1). To do this, we take any solution $y$ of equation (1) and consider the difference $y - \tilde{y}(x)$. It is the solution of the homogeneous equation (2). Indeed,

$$[y - \tilde{y}(x)]'' + p(x)\,[y - \tilde{y}(x)]' + q(x)\,[y - \tilde{y}(x)]$$
$$= [y'' + p(x)\,y' + q(x)\,y] - [\tilde{y}''(x) + p(x)\,\tilde{y}'(x)$$
$$+ q(x)\,\tilde{y}(x)] = f(x) - f(x) = 0.$$

This means that the difference $y - \tilde{y}(x)$ can be written in the form

$$y - \tilde{y}(x)\,C_1^0y_1(x) + C_2^0y_2(x),$$

whence

$$y = \tilde{y}(x) + C_1^0y_1(x) + C_2^0y_2(x),$$

where $C_1^0$ and $C_2^0$ are definite values of the constants $C_1$ and $C_2$. Thus any solution $y$ of equation (1) results from formula (10) for an appropriate choice of the arbitrary constants $C_1$ and $C_2$, i.e. function (10) is the general solution of equation (1). ■

Thus, to find the general solution of a nonhomogeneous linear equation, we must find the general solution of the corresponding homogeneous equation and some particular solution of the nonhomogeneous equation. In the general case the problem of finding a particular solution is complicated. We shall show how to find a particular solution of a nonhomogeneous equation by the method of variation of arbitrary constants when the general solution of the corresponding homogeneous equation is known.

Let $Y = C_1y_1(x) + C_2y_2(x)$ be the general solution of the homogeneous equation (2). We shall seek a particular solution of the nonhomogeneous equation (1) in the form

$$y = C_1(x)\,y_1(x) + C_2(x)\,y_2(x), \tag{11}$$

considering $C_1$ and $C_2$ to be some required functions of $x$. We differentiate the last equality

$$y' = C_1'(x)\,y_1(x) + C_1(x)\,y_1'(x) + C_2'(x)\,y_2(x) + C_2(x)\,y_2'(x). \tag{12}$$

We choose functions $C_1(x)$ and $C_2(x)$ such that the equality

$$C_1'(x)\,y_1(x) + C_2'(x)\,y_2(x) = 0 \tag{13}$$

holds true. Then equality (12) takes the form

$$y' = C_1(x)\,y_1'(x) + C_2(x)\,y_2'(x).$$

Differentiating this equality, we find $y''$:

$$y'' = C_1'(x)\,y_1'(x) + C_1(x)\,y_1''(x) + C_2'(x)\,y_2'(x) + C_2(x)\,y_2''(x).$$

Substituting the expressions for $y$, $y'$, $y''$ into equation (1) and grouping terms, we obtain

$$C_1(x)\,[y_1''(x) + p(x)\,y_1'(x) + q(x)\,y_1(x)]$$
$$+ C_2(x)\,[y_2''(x) + p(x)\,y_2'(x) + q(x)\,y_2(x)] + C_1'(x)\,y_1'(x)$$
$$+ C_2'(x)\,y_2'(x) = f(x).$$

The expressions in brackets are zero since $y_1(x)$ and $y_2(x)$ are solutions of the homogeneous equation. Therefore the last equality assumes the form

$$C_1'(x)\,y_1'(x) + C_2'(x)\,y_2'(x) = f(x). \tag{14}$$

Thus function (11) is a solution of equation (1) if the functions $C_1(x)$ and $C_2(x)$ satisfy equalities (13) and (14). Uniting them, we

get a system of equations

$$\begin{cases} C_1'(x)y_1(x) + C_2'(x)\,y_2(x) = 0, \\ C_1'(x)y_1'(x) + C_2'(x)\,y_2'(x) = f(x) \end{cases} \tag{15}$$

in which $C_1'(x)$ and $C_2'(x)$ are unknown and $y_1(x)$, $y_2(x)$, $y_1'(x)$, $y_2'(x)$, $f(x)$ are known. Since the determinant of this system is the Wronskian

$$W(x) = \begin{vmatrix} y_1(x) & y_2(x) \\ y_1'(x) & y_2'(x) \end{vmatrix}$$

composed of the linearly independent solutions $y_1(x)$ and $y_2(x)$ of the homogeneous equation (2), it follows, according to Theorem 15.5, that it is nonzero and, hence, system (15) has a unique solution with respect to $C_1'(x)$ and $C_2'(x)$. Solving this system, we obtain

$$C_1'(x) = \varphi_1(x), \quad C_2'(x) = \varphi_2(x),$$

where $\varphi_1(x)$ and $\varphi_2(x)$ are known functions, whence, by means of integration, we find $C_1(x)$ and $C_2(x)$. Substituting the expressions obtained for $C_1(x)$ and $C_2(x)$ into equality (11), we get the required particular solution of equation (1).

**Example 2.** Find a particular solution of the equation $y'' - y = x$.

*Solution.* In Example 1, Item 2, we found the general solution $Y(x) = C_1e^x + C_2e^{-x}$ of the corresponding homogeneous equation $y'' - y = 0$. Therefore we shall seek a particular solution of the nonhomogeneous equation in the form

$$\tilde{y}(x) = C_1(x)\,e^x + C_2(x)\,e^{-x}. \tag{16}$$

In this case system (15) for finding $C_1'(x)$ and $C_2'(x)$ has the form

$$\begin{cases} C_1'(x)\,e^x + C_2'(x)\,e^{-x} = 0, \\ C_1'(x)\,e^x - C_2'(x)\,e^{-x} = x. \end{cases}$$

Adding these equations together, we find that $C_1'(x) = (1/2)xe^{-x}$. By means of integration, we find from this that

$$C_1(x) = -\frac{1}{2}(x+1)\,e^{-x}.$$

We do not write the arbitrary constant since we seek a particular solution. Substituting the expession for $C_1'(x)$ into the first equation of the system, we find that $C_2'(x) = -(1/2)\,xe^x$, and from this, by means of integration, we find that

$$C_2(x) = -\frac{1}{2}(x-1)\,e^x.$$

Substituting the expressions obtained for $C_1(x)$ and $C_2(x)$ into equality (16), we get a particular solution $\tilde{y}$ of the given nonhomogeneous equation:

$$\tilde{y}(x) = \left[-\frac{1}{2}(x+1)e^{-x}\right]e^{x} + \left[-\frac{1}{2}(x-1)e^{x}\right]e^{-x} = -x.$$

Note that having found a particular solution of the nonhomogeneous equation and knowing the general solution of the corresponding homogeneous equation, we can write, by virtue of (10), the general solution of the given nonhomogeneous equation:

$$y = \tilde{y}(x) + Y(x) = -x + C_1e^{x} + C_2e^{-x},$$

where $C_1$ and $C_2$ are arbitrary constants.

### 3.15.4. Second-Order Linear Differential Equations with Constant Coefficients

We shall discuss an important particular case of second-order linear differential equations, the case when the function $p(x)$ and $q(x)$ are constant quantities. Equations of this kind are known as linear equations with constant coefficients.

**1. Second-order homogeneous linear differential equations with constant coefficients.** Let us consider a second-order homogeneous linear equation

$$y'' + py' + qy = 0, \tag{1}$$

where $p$ and $q$ are real numbers.

**Theorem 15.8.** (1) *If the number $k$ is a real root of the equation*

$$k^2 + pk + q = 0, \tag{2}$$

*then the function $y = e^{kx}$ is a solution of equation* (1).

(2) *If the numbers $k_1 = \alpha + i\beta$ and $k_2 = \alpha - i\beta$ $(\beta \neq 0)$ are complex roots of equation* (2), *then the functions $y_1 = e^{\alpha x}\cos\beta x$ and $y_2 = e^{\alpha x}\sin\beta x$ are solutions of equation* (1).

*Proof.* (1) Let $y = e^{kx}$ ($k = \text{const}$). Then $y' = ke^{kx}$ and $y'' = k^2e^{kx}$. Substituting $y$, $y'$, $y''$ into equation (1), we get

$$e^{kx}(k^2 + pk + q) = 0.$$

Since $e^{kx} \neq 0$, we can cancel $e^{kx}$ and get

$$k^2 + pk + q = 0.$$

Consequently, if $k$ is a root of equation (2), then the function $y = e^{kx}$ is a solution of equation (1).

(2) By analogy with case (1) we can verify that the functions $y_1 = e^{\alpha x}\cos\beta x$ and $y_2 = e^{\alpha x}\sin\beta x$ satisfy equation (1). ■

Equation (2) is the *characteristic equation* of the given equation (1).

The characteristic equation (2) is a quadratic equation and, consequently, has two roots. We designate them as $k_1$, and $k_2$.

**Theorem 15.9.** (1) *If the roots of the characteristic equation are real and distinct* ($k_1 \neq k_2$), *then the general solution of equation* (1) *has the form*

$$y = C_1 e^{k_1 x} + C_2 e^{k_2 x};$$

(2) *if the roots of the characteristic equation are real and equal* ($k_1 = k_2$), *then the general solution has the form*

$$y = C_1 e^{k_1 x} + C_2 x e^{k_1 x};$$

(3) *if the roots of the characteristic equation are complex* ($k_1 = \alpha + i\beta$ *and* $k_2 = \alpha - i\beta$, $\beta \neq 0$), *then the general solution has the form*

$$y = e^{\alpha x}(C_1 \cos \beta x + C_2 \sin \beta x).$$

*Proof.* (1) Let the roots $k_1$ and $k_2$ be distinct. According to Theorem 15.8, the functions $y_1 = e^{k_1 x}$ and $y_2 = e^{k_2 x}$ are particular solutions of equation (1). These solutions are linearly independent since $y_2/y_1 = e^{(k_2 - k_1)x} \neq \text{const}$ ($k_1 \neq k_2$). Consequently, according to Theorem 15.6, the general solution of equation (1) has the form

$$y = C_1 e^{k_1 x} + C_2 e^{k_2 x}.$$

(2) Let the roots $k_1$ and $k_2$ be equal, i.e. $k_1 = k_2$. According to Theorem 15.8, the function $y_1 = e^{k_1 x}$ is a particular solution of equation (1). Let us find the second particular solution which is linearly independent of the first. We shall seek it in the form $y_2 = z(x)e^{k_1 x}$, where $z(x)$ is a new unknown function. Differentiating, we obtain

$$y_2' = z' e^{k_1 x} + k_1 z e^{k_1 x} = e^{k_1 x}(z' + k_1 z),$$

$$y_2'' = z'' e^{k_1 x} + 2k_1 z' e^{k_1 x} + k_1^2 z e^{k_1 x} = e^{k_1 x}(z'' + 2k_1 z' + k_1^2 z).$$

Substituting $y_2$, $y_2'$, $y_2''$ into the left-hand side of equation (1), we get

$$e^{k_1 x}\left[z'' + 2\left(k_1 + \frac{p}{2}\right) z' + (k_1^2 + pk_1 + q)\, z\right] = 0.$$

By the hypothesis, $k_1^2 + pk_1 + q = 0$. In addition, $k_1 = k_2 = -p/2$ and therefore $k_1 + p/2 = 0$. Consequently, to find the function $z(x)$, we must solve the equation $z'' = 0$. Carrying out successive integration, we get $z' = \tilde{C}_1$ and $z = \tilde{C}_1 x + \tilde{C}_2$, where $\tilde{C}_1$ and $\tilde{C}_2$ are arbitrary constants. Setting $\tilde{C}_1 = 1$ and $\tilde{C}_2 = 0$, we find that $z(x) = x$. Thus $y_2 = xe^{k_1 x}$ is the second particular solution of equation (1). The solutions $y_1$ and $y_2$ are linearly independent since $y_1/y_2 = 1/x \neq \text{const}$, and, according to Theorem 15.6, the general

solution of equation (1) has the form

$$y = C_1 e^{k_1 x} + C_2 x e^{k_1 x}.$$

(3) Let $k_1$ and $k_2$ be complex conjugate roots, i.e. $k_1 = \alpha + i\beta$ and $k_2 = \alpha - i\beta$ ($\beta \neq 0$). Then, according to Theorem 15.8, the functions $y_1 = e^{\alpha x} \cos \beta x$ and $y_2 = e^{\alpha x} \sin \beta x$ are particular solutions of equation (1). These solutions are linearly independent since $y_1/y_2 = \cot \beta x \neq \text{const}$. Therefore the general solution of equation (1) has the form

$$y = e^{\alpha x}(C_1 \cos \beta x + C_2 \sin \beta x). \blacksquare$$

**Example 1.** Find the general solution of the equation $y'' + y' - 2y = 0$.

*Solution.* The characteristic equation has the form $k^2 + k - 2 = 0$, its roots $k_1 = 1$ and $k_2 = -2$ being real and distinct. The corresponding particular solutions of the equation are $y_1 = e^x$ and $y_2 = e^{-2x}$. The general solution of the equation has the form

$$y = C_1 e^x + C_2 e^{-2x}.$$

**Example 2.** Find the general solution of the equation $y'' - 2y' + y = 0$.

*Solution.* The characteristic equation has the form $k^2 - 2k + 1 = 0$, its roots $k_1 = k_2 = 1$ being real and equal. The corresponding particular solutions of the equation are $y_1 = e^x$ and $y_2 = xe^x$. The general solution of the equation has the form

$$y = C_1 e^x + C_2 x e^x = e^x (C_1 + C_2 x).$$

**Example 3.** Find the general solution of the equation $y'' - 4y' + 13y = 0$.

*Solution.* The characteristic equation has the form $k^2 - 4k + 13 = 0$, its roots $k_1 = 2 + i3$ and $k_2 = 2 - i3$ being complex. The corresponding particular solutions are $y_1 = e^{2x} \cos 3x$ and $y_2 = e^{2x} \sin 3x$. The general solution of the equation has the form

$$y = e^{2x} (C_1 \cos 3x + C_2 \sin 3x).$$

**2. Second-order nonhomogeneous linear differential equations with constant coefficients.** Consider a second-order nonhomogeneous linear equation

$$y'' + py' + qy = f(x), \tag{3}$$

where $p$ and $q$ are real numbers and $f(x)$ is a continuous function.

We know that the general solution of such an equation is the sum of a particular solution of a nonhomogeneous equation and the general solution of the corresponding homogeneous equation. We know how to find the general solution of a homogeneous equation and so it remains to discuss the problem of finding a particular solution. To find a particular solution, we can use the method of

variation of arbitrary constants. However, if we have a polynomial, or an exponential function, or a trigonometric function $\sin \beta x$ or $\cos \beta x$, or, else, a linear combination of the enumerated functions on the right-hand side of equation (3), then we can find a particular solution by the method of undetermined coefficients which does not include the process of integration.

Let us consider various kinds of right-hand sides of equation (3).

(1) *The right-hand side has the form*

$$f(x) = P_n(x),$$

*where* $P_n(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n$ *is a polynomial of degree n.* Then we can seek the particular solution in the form

$$\tilde{y} = Q_n(x)\, x^r,$$

where $Q_n(x)$ is a polynomial of the same degree as $P_n(x)$ and $r$ is the number of roots of the characteristic equation which are equal to zero.

**Example 4.** Find the general solution of the equation $y'' - 2y' + y = x + 1$.

*Solution.* The general solution of the corresponding homogeneous equation has the form $Y = (C_1 + C_2x)\, e^x$ (see Example 2). Since the right-hand side of the equation is a polynomial of the first degree and none of the roots of the characteristic equation $k^2 - 2k + 1 = 0$ is zero ($k_2 = k_1 = 1$), we seek the particular solution in the form

$$\tilde{y} = (Ax + B)\, x^0 = Ax + B,$$

where $A$ and $B$ are unknown coefficients. Differentiating $\tilde{y} = Ax + B$ twice and substituting $\tilde{y}$, $\tilde{y}'$, $\tilde{y}''$ into the given equation, we find that

$$-2A + Ax + B = x + 1.$$

Equating the coefficients in the like powers of $x$ on both sides of the relation, i.e. $A = 1$ and $-2A + B = 1$, we find that $A = 1$ and $B = 3$. Thus the particular solution of the given equation has the form $\tilde{y} = x + 3$ and the general solution is

$$y = (C_1 + C_2x)\, e^x + (x + 3).$$

(2) *The right-hand side has the form*

$$f(x) = e^{\alpha x} P_n(x),$$

*where* $P_n(x)$ *is a polynomial of degree n.* Then we must seek the particular solution in the form

$$\tilde{y} = Q_n(x)\, x^r e^{\alpha x},$$

where $Q_n(x)$ is a polynomial of the same degree as $P_n(x)$ and $r$ is the number of roots of the characteristic equation which are equal to $\alpha$. If $\alpha = 0$, then $f(x) = P_n(x)$, i.e. we have case (1).

**Example 5.** Find the general solution of the equation $y'' - 4y' + 3y = xe^x$.

*Solution.* The characteristic equation $k^2 - 4k + 3 = 0$ has roots $k_1 = 1$ and $k_2 = 3$. This means that the general solution of the corresponding homogeneous equation has the form $Y = C_1e^x + C_2e^{3x}$. The right-hand side of this equation includes the product of a first-degree polynomial by the exponential function $e^{\alpha x}$ for $\alpha = 1$. Since there is only one root $k_1 = \alpha = 1$ among the roots of the characteristic equation, it follows that $r = 1$. In this case $P_n(x) = x$ is a first-degree polynomial. Therefore we seek the particular solution of the given equation in the form

$$\tilde{y} = (Ax + B)\,xe^x = (Ax^2 + Bx)\,e^x.$$

Differentiating $\tilde{y} = (Ax + B)xe^x$ twice and substituting $y$, $y'$, $y''$ into the equation, we get

$$-4Ax + 2A - 2B = x.$$

Equating the coefficients in the like powers of $x$ on both sides of the relation, i.e. $-4A = 1$ and $2A - 2B = 0$, we find that $A = -1/4$ and $B = -1/4$. Substituting the values obtained for $A$ and $B$ into the expression for $\tilde{y}$, we get the particular solution of the given equation $\tilde{y} = -(1/4)\,(x^2 + x)\,e^x$. The general solution has the form

$$y = \tilde{y} + Y = C_1e^x + C_2e^{3x} - \frac{1}{4}(x^2 + x)\,e^x.$$

(3) *The right-hand side has the form*

$$f(x) = a\cos\beta x + b\sin\beta x,$$

*where $a$, $b$, $\beta$ are known numbers.* Then we must seek the particular solution in the form

$$\tilde{y} = (A\cos\beta x + B\sin\beta x)\,x^r,$$

where $A$ and $B$ are unknown coefficients and $r$ is the number of roots of the characteristic equation which are equal to $i\beta$.

**Example 6.** Find the general solution of the equation $y'' + y = \sin x$.

*Solution.* The characteristic equation $k^2 + 1 = 0$ has roots $k_1 = i$ and $k_2 = -i$. Therefore the general solution of the corresponding homogeneous equation is $Y = C_1\cos x + C_2\sin x$. The right-hand side of the relation includes a trigonometric function $\sin x$, i.e. $a =$

0, $b = 1$, $\beta = 1$. Since $i\beta = i$ is a root of the characteristic equation, it follows that $r = 1$ and we must seek the particular solution in the form

$$\tilde{y} = (A \cos x + B \sin x)\, x.$$

Differentiating $\tilde{y} = (A \cos x + B \sin x)\, x$ twice and substituting $y$, $y'$, $y''$ into the equation, we get

$$2\,(-A \sin x + B \cos x) = \sin x,$$

whence $A = -1/2$ and $B = 0$. Thus the particular solution is $\tilde{y} = -(1/2)\, x \cos x$ and the general solution is

$$y = \tilde{y} + Y = C_1 \cos x + C_2 \sin x - \frac{1}{2}\, x \cos x.$$

**Example 7.** Find the general solution of the equation $y'' + y = \sin 2x$.

*Solution.* This equation differs from the preceding one only by $\beta = 2$. Since $i\beta = i2$ is not a root of the characteristic equation, it follows that $r = 0$ and we must seek the particular solution in the form

$$\tilde{y} = A \cos 2x + B \sin 2x.$$

Differentiating $\tilde{y} = A \cos 2x + B \sin 2x$ twice and substituting $y$, $y'$, $y''$ into the equation, we get

$$-3A \cos 2x - 3B \sin 2x = \sin 2x,$$

whence $A = 0$ and $B = -1/3$, i.e. the particular solution is $\tilde{y} = -(1/3) \sin 2x$ and the general solution is

$$y = \tilde{y} + Y = C_1 \cos x + C_2 \sin x - \frac{1}{3} \sin 2x.$$

(4) *The right-hand side has the form*

$$f(x) = e^{\alpha x}[P_n(x) \cos \beta x + P_m(x) \sin \beta x],$$

*where $P_n(x)$ is a polynomial of degree $n$ and $P_m(x)$ is a polynomial of degree $m$.* Then we must seek the particular solution in the form

$$\tilde{y} = x^r e^{\alpha x}\, [Q_1(x) \cos \beta x + Q_2(x) \sin \beta x],$$

where $Q_1(x)$ and $Q_2(x)$ are polynomials of degree $s$, $s = \max\{n, m\}$ and $r$ is the number of roots of the characteristic equation which are equal to $\alpha + i\beta$.

**Example 8.** Find the general solution of the equation $y'' - y = 3e^{2x} \cos x$.

*Solution.* The characteristic equation $k^2 - 1 = 0$ has roots $k_1 = 1$ and $k_2 = -1$. The general solution of the homogeneous equa-

tion is $Y = C_1e^x + C_2e^{-x}$. The right-hand side includes the product of a zero-degree polynomial, an exponential and a trigonometric function so that $P_n(x) = 3$, $P_m(x) = 0$, $s = 0$. The number $\alpha + i\beta = 2 + i1$ is not a root of the characteristic equation and therefore $r = 0$, and we seek the particular solution in the form

$$\tilde{y} = e^{2x}(A \cos x + B \sin x).$$

Differentiating $\tilde{y} = e^{2x}(A \cos x + B \sin x)$ twice and substituting $y$, $y'$, $y''$ into the equation, we get

$$(2A + 4B) \cos x + (2B - 4A) \sin x = 3 \cos x.$$

Equating the coefficients in $\cos x$ and $\sin x$, we obtain

$$2A + 4B = 3, \quad -4A + 2B = 0,$$

whence $A = 3/10$ and $B = 3/5$. Thus the particular solution is $\tilde{y} = e^{2x}\left(\frac{3}{10} \cos x + \frac{3}{5} \sin x\right)$ and the general solution of the equation is

$$y = \tilde{y} + Y = e^{2x}\left(\frac{3}{10} \cos x + \frac{3}{5} \sin x\right) + C_1e^x + C_2e^{-x}.$$

**Example 9.** Proceeding from the given roots of the characteristic equation and the right-hand side of $f(x)$, write the particular solution $\tilde{y}$ of each of the following nonhomogeneous linear equations:

(a) $k_1 = 3 + i2$, $k_2 = 3 - i2$, $f(x) = 8e^{3x} \sin 2x$;

(b) $k_1 = k_2 = -3$, $f(x) = 2xe^{-3x} \sin x$;

(c) $k_1 = 1$, $k_2 = -3$, $f(x) = e^x(1 - x) \cos 3x$;

(d) $k_1 = 1 + i2$, $k_2 = 1 - i2$, $f(x) = e^x(\cos 2x - 3 \sin 2x)$;

(e) $k_1 = 2 + i\frac{1}{2}$, $k_2 = 2 - i\frac{1}{2}$, $f(x) = e^{2x}\left[(x^3 - 1)\cos\frac{x}{2} + x \sin\frac{x}{2}\right]$.

*Solution.* (a) We have $\alpha = 3$, $\beta = 2$, $P_n(x) = 0$, $P_m(x) = 8$, $s = 0$. Since the number $\alpha + i\beta = 3 + i2$ is a root of the characteristic equation, it follows that $r = 1$. Therefore $\tilde{y} = xe^{3x}(A \cos 2x + B \sin 2x)$.

(b) We have $\alpha = -3$, $\beta = 1$, $P_n(x) = 0$, $P_m(x) = 2x$, $m = 1$, $s = 1$. The number $\alpha + i\beta = -3 + i$ is not a root of the characteristic equation and therefore $r = 0$. Consequently, $\tilde{y} = e^{-3x}[(Ax + B) \cos x + (Cx + D) \sin x]$.

(c) We have $\alpha = 1$, $\beta = 3$, $P_n(x) = 1 - x$, $P_m(x) = 0$, $n = 1$, $s = 1$. Since the number $\alpha + i\beta = 1 + i3$ is not a root of the characteristic equation, it follows that $r = 0$. Therefore $\tilde{y} = e^x[(Ax + B) \cos 3x + (Cx + D) \sin 3x]$.

(d) We have $\alpha = 1$, $\beta = 2$, $P_n(x) = 1$, $P_m(x) = -3$, $s = 0$. The number $\alpha + i\beta = 1 + i2$ is a root of the characteristic equation and therefore $r = 1$. Consequently, $\tilde{y} = xe^x(A\cos 2x + B\sin 2x)$.

(e) We have $\alpha = 2$, $\beta = 1/2$, $P_n(x) = (x^3 - 1)$, $P_m(x) = x$, $n = 3$, $m = 1$, $s = 3$. The number $\alpha + i\beta = 2 + i\frac{1}{2}$ is a root of the characteristic equation, and, hence, $r = 1$. Consequently, $\tilde{y} = xe^{2x}[(Ax^3 + Bx^2 + Cx + D)\cos(x/2) + (Ex^3 + Fx^2 + Gx + H)\sin(x/2)]$.

In conclusion, we shall prove a theorem which is often used in solving nonhomogeneous linear equations whose right-hand side includes a sum of several terms.

**Theorem 15.10.** *If $\tilde{y}_1$ is a solution of the equation*

$$y'' + py' + qy = f_1(x) \tag{4}$$

*and $\tilde{y}_2$ is a solution of the equation*

$$y'' + py' + qy = f_2(x), \tag{5}$$

*then the sum $\tilde{y}_1 + \tilde{y}_2$ is a solution of the equation*

$$y'' + py' + qy = f_1(x) + f_2(x). \tag{6}$$

*Proof.* We set up a sum $\tilde{y}_1 + \tilde{y}_2$ and substitute it into the left-hand side of equation (6). Then we obtain

$$(\tilde{y}_1 + \tilde{y}_2)'' + p(\tilde{y}_1 + \tilde{y}_2)' + q(\tilde{y}_1 + \tilde{y}_2)$$
$$= (\tilde{y}_1'' + p\tilde{y}_1' + q\tilde{y}_1) + (\tilde{y}_2'' + p\tilde{y}_2' + q\tilde{y}_2) = f_1 + f_2$$

since, by the hypothesis, the expression in the first parentheses is equal to $f_1(x)$ and that in the second parentheses is equal to $f_2(x)$. Consequently, $\tilde{y}_1 + \tilde{y}_2$ is a solution of equation (6). ■

**Example 10.** Find the general solution of the equation $y'' - 2y' + y = \sin x + e^{-x}$.

*Solution.* The characteristic equation $k^2 - 2k + 1 = 0$ has roots $k_1 = k_2 = 1$ and therefore the general solution of the corresponding homogeneous equation is $Y = C_1e^x + C_2xe^x = e^x(C_1 + C_2x)$.

Since the right-hand side of the equation consists of the sum of two functions $\sin x$ and $e^{-x}$, it follows, according to Theorem 15.10, that the particular solution of the given equation must be sought in the form

$$\tilde{y} = \tilde{y}_1 + \tilde{y}_2,$$

where $\tilde{y}_1$ is a particular solution of the equation $y'' - 2y' + y = \sin x$ and $\tilde{y}_2$ is a particular solution of the equation $y'' - 2y' + y = e^{-x}$.

We shall first find the particular solution $\tilde{y}_1$. Since the number $i\beta = i$ is not a root of the characteristic equation ($r = 0$), we shall seek the particular solution in the form $\tilde{y}_1 = A \sin x + B \cos x$. Substituting $\tilde{y}_1$, $\tilde{y}_1'$, $\tilde{y}_1''$ into the equation $y'' - 2y' + y = \sin x$ and equating the coefficients in $\sin x$ and $\cos x$, we obtain $-2A = 0$ and $2B = 1$, whence $A = 0$, $B = 1/2$ and, consequently, $\tilde{y}_1 = (1/2) \cos x$.

We shall find now the particular solution $\tilde{y}_2$. We seek it in the form $\tilde{y}_2 = Ae^{-x}$ since the number $\alpha = -1$ is not a root of the characteristic equation. Substituting $\tilde{y}_2$, $\tilde{y}_2'$, $\tilde{y}_2''$ into the equation $y'' - 2y' + y = e^{-x}$, we get $A = 1/4$. Consequently, $\tilde{y}_2 = (1/4)e^{-x}$. Thus the particular solution of the given equation has the form

$$\tilde{y} = \tilde{y}_1 + \tilde{y}_2 = \frac{1}{2} \cos x + \frac{1}{4} e^{-x}$$

and the general solution of this equation is

$$y = \tilde{y} + Y = \frac{1}{2} \cos x + \frac{1}{4} e^{-x} + C_1 e^x + C_2 x e^x.$$

### 3.15.5. Employing Linear Differential Equations When Studying Oscillatory Motions

Linear differential equations constitute a powerful apparatus used in the solution of problems on oscillations which are very important for modern physics and engineering. We shall get acquainted with one of them, the problem on the oscillation of a load suspended by a vertical spring.

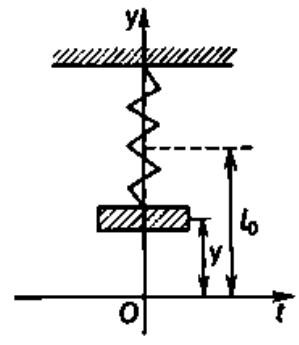


Fig. 227

**Statement of the problem.** Let a load of mass $m$ suspended by a spring move along a vertical straight line. If we extend or contract the spring with the load, the load will begin oscillating about the position of equilibrium. Let us establish the law of motion of the load, i.e. find a formula expressing the deviation of the load from the position of equilibrium at any moment $t$.

We make the origin coincident with the equilibrium position of the load and direct the $y$-axis vertically upwards. We designate the distance from the end of the nonextended spring without a load to the position of equilibrium of the load as $l_0$ and the deviation of the load from the equilibrium position at the moment $t$ as $y$ (Fig. 227). The force that acts on the load is equal to the sum of the following

three forces: (1) the gravity $mg$ of the load which is directed downwards; (2) the resistance of the medium which is directed opposite to the motion of the load and is proportional in value to the speed of motion of the load, i.e. is equal to $\lambda \frac{dy}{dt}$, where $\lambda$ is the proportionality factor; (3) the elastic force of the spring which is directed upwards (i.e. in the positive direction of the $y$-axis) and whose value, according to Hooke's* law, is proportional to the deformation, i.e. is equal to $c(l_0 - y)$, where $c$ is the proportionality factor known as the coefficient of stiffness of the spring (the mass of the spring is neglected).

According to Newton's second law, we get the following equation for the motion of the load:

$$m\frac{d^2y}{dt^2} = c(l_0 - y) - mg - \lambda\frac{dy}{dt}.$$

Since in the position of equilibrium ($y = 0$) the weight $mg$ of the load is balanced by the elastic force of the spring, we have $mg = cl_0$, and therefore

$$m\frac{d^2y}{dt^2} = -cy - \lambda\frac{dy}{dt},$$

or

$$m\frac{d^2y}{dt^2} + \lambda\frac{dy}{dt} + cy = 0. \tag{1}$$

We have obtained a differential equation which is known as the *equation of free oscillations of a load suspended by a spring.*

If a load is acted upon by an external force directed vertically (along the $y$-axis), whose value $F(t)$ depends on time $t$, then equation (1) assumes the form

$$m\frac{d^2y}{dt^2} + \lambda\frac{dy}{dt} + cy = F(t). \tag{2}$$

Equation (2) is known as the *equation of forced oscillations of a load suspended by a spring.*

Dividing all the terms of equation (2) by $m$ and designating $c/m = \omega^2$, $\lambda/m = 2\mu$, $F(t)/m = f(t)$, we get the final form of the equation of forced oscillations:

$$\frac{d^2y}{dt^2} + 2\mu\frac{dy}{dt} + \omega^2 y = f(t). \tag{3}$$

Equation (3) is a *second-order nonhomogeneous linear differential equation with constant coefficients.*

Let us consider now oscillations using the familiar solutions of linear equations with constant coefficients.

* Hooke, Robert (1635-1703), an English physicist and natural philosopher.

**Free oscillations.** Assume that the external force $f(t)$ and the resistance $\mu$ of the medium are absent. Then equation (3) assumes the form

$$\frac{d^2y}{dt^2} + \omega^2 y = 0.$$

This is a homogeneous linear equation. The characteristic equation $k^2 + \omega^2 = 0$ has roots $k_1 = i\omega$ and $k_2 = -i\omega$. The general solution of the equation is defined by the formula

$$y = C_1 \cos \omega t + C_2 \sin \omega t,$$

where $C_1$ and $C_2$ are arbitrary constants. To make further arguments more convenient, we replace the arbitrary constants $C_1$ and $C_2$ by the constants $A > 0$ and $\varphi$, setting $C_1 = A \sin \varphi$ and $C_2 = A \cos \varphi$ (hence $A = \sqrt{C_1^2 + C_2^2}$ and $\tan \varphi = C_1/C_2$). Then

$$y = A \sin \varphi \cos \omega t + A \cos \varphi \sin \omega t$$

and we can write the general solution as

$$y = A \sin (\omega t + \varphi). \tag{4}$$

Formula (4) expresses the law of motion of the load suspended by a spring, i.e. the deviation $y$ of the load from the position of equilibrium at any moment $t$. According to this formula, the load makes, as we say, free harmonic oscillations about the equilibrium position. The quantity $A$ is the *amplitude of the oscillations*, $\omega$ is the *frequency of the oscillations* and $\varphi$ is the *initial phase*.

To isolate a particular solution from the general one, we must specify the initial conditions of the motion. Let, at the initial moment $t = 0$, the deviation and the speed of the load be known:

$$y|_{t=0} = y_0, \quad y'|_{t=0} = y_0'. \tag{5}$$

Then differentiation results in

$$y' = A\omega \cos (\omega t + \varphi). \tag{6}$$

Substituting the initial conditions (5) into (4) and (6), we obtain

$$\begin{cases} y_0 = A \sin \varphi, \\ y_0' = A\omega \cos \varphi. \end{cases} \tag{7}$$

Expressing the arbitrary constants $A$ and $\varphi$ in terms of $\omega$, $y_0$, $y_0'$ and substituting their values into (4), we get from this the required particular solution which satisfies the initial conditions (5).

It follows, in particular, from formulas (7) that the constants $A$ and $\varphi$ depend on the frequency $\omega$ of oscillations and the initial conditions of the motion. As to the frequency of oscillations, it does not depend on the initial conditions but depends on the ratio of the coefficient of stiffness of the spring to the mass of the load ($\omega = \sqrt{c/m}$).

Assume now that the external force $f(t)$ is absent but there is resistance of the medium ($\mu \neq 0$), say, the air drag. In this case equation (3) assumes the form

$$\frac{d^2y}{dt^2} + 2\mu\frac{dy}{dt} + \omega^2 y = 0. \tag{8}$$

The characteristic equation $k^2 + 2\mu k + \omega^2 = 0$ has roots $k_{1,2} = -\mu \pm \sqrt{\mu^2 - \omega^2}$. Three cases are possible here.

(1) $\mu > \omega$. Then the roots $k_1 = -\mu + \sqrt{\mu^2 - \omega^2}$ and $k_2 = -\mu - \sqrt{\mu^2 - \omega^2}$ are real, distinct and negative. The general solution of equation (8) has the form

$$y = C_1 e^{(-\mu+\sqrt{\mu^2-\omega^2})t} + C_2 e^{(-\mu-\sqrt{\mu^2-\omega^2})t}.$$

It follows from the solution obtained that the load performs no oscillations, and with an infinite increase in $t$ the deviation $y$ of the load infinitely slowly approaches the equilibrium position ($y \to 0$ as $t \to +\infty$). Then we say that the load performs a nonperiodic damped motion.

(2) $\mu = \omega$. Then the roots $k_1 = k_2 = -\mu$ are real, equal and negative. The general solution of equation (8) has the form

$$y = C_1 e^{-\mu t} + C_2 t e^{-\mu t} = e^{-\mu t}(C_1 + C_2 t).$$

In this case the load performs a motion similar to the preceding one.

(3) $\mu < \omega$. Then the roots $k_1 = -\mu + i\sqrt{\omega^2 - \mu^2}$ and $k_2 = -\mu - i\sqrt{\omega^2 - \mu^2}$ are complex. The general solution of equation (8) has the form

$$y = e^{-\mu t}(C_1 \cos \tilde{\omega} t + C_2 \sin \tilde{\omega} t),$$

where $\tilde{\omega} = \sqrt{\omega^2 - \mu^2}$. Replacing $C_1$ and $C_2$ by the constants $A$ and $\varphi$, we write the solution of equation (8) in the form

$$y = A e^{-\mu t} \sin (\tilde{\omega} t + \varphi).$$

As distinct from formula (4), the amplitude $Ae^{-\mu t}$ here depends on time $t$. Since $\mu < 0$, the amplitude tends to zero as $t \to +\infty$. Therefore, in this case, the load performs free damped oscillations about the equilibrium position.

**Forced oscillations. Resonance.** Let us consider now the case when an oscillating system is acted upon by a periodic external force $f(t) = a \sin \omega_1 t$, assuming for simplicity that there is no resistance of the medium ($\mu = 0$). Then equation (8) assumes the form

$$d^2y/dt^2 + \omega^2 y = a \sin \omega_1 t. \tag{9}$$

This is a nonhomogeneous linear equation with constant coefficients. We know that the general solution of this equation is the sum of the

general solution $Y$ of the corresponding homogeneous equation which we have found earlier (see formula (4)) and the particular solution $\tilde{y}$ of the nonhomogeneous equation which we have to find.

We shall consider two cases separately.

(a) $\omega \neq \omega_1$, i.e. the frequency of the periodic external force is different from the frequency of the free oscillations of the load. Since the number $i\omega_1$ does not coincide with the root of the characteristic equation $k^2 + \omega^2 = 0$, we can find the particular solution in the form

$$\tilde{y} = A \cos \omega_1 t + B \sin \omega_1 t.$$

Differentiating $\tilde{y}$ twice and substituting $\tilde{y}$ and $\tilde{y}''$ into equation (9), we find that $B = a/(\omega^2 - \omega_1^2)$ and $A = 0$. Thus

$$\tilde{y} = \frac{a}{\omega^2 - \omega_1^2} \sin \omega_1 t$$

and the general solution of equation (9) has the form

$$y = \tilde{y} + Y = \frac{a}{\omega^2 - \omega_1^2} \sin \omega_1 t + A \sin (\omega t + \varphi). \tag{10}$$

As follows from formula (10), the particular solution $\tilde{y}$ defines the oscillation of the system produced by the external force, the general solution $Y = A \sin (\omega t + \varphi)$ is the free oscillation of the load, and the general solution $y$ is a composite oscillatory motion resulting from the addition of two oscillations with different frequencies $\omega$ and $\omega_1$.

In this case the amplitude is constant, and if $\omega$ and $\omega_1$ are close in value, then the load oscillates about the position of equilibrium with large amplitude.

(b) $\omega = \omega_1$, i.e. the frequency of the periodic external force coincides with the frequency of the free oscillations of the load. Since $i\omega_1 = i\omega$ is a root of the characteristic equation $k^2 + \omega^2 = 0$, we must seek the particular solution in the form

$$\tilde{y} = t (A \cos \omega t + B \sin \omega t).$$

Differentiating $\tilde{y}$ twice and substituting $\tilde{y}$ and $\tilde{y}''$ into equation (9), we find that $A = -a/(2\omega)$ and $B = 0$. Thus

$$\tilde{y} = -\frac{at}{2\omega} \cos \omega t$$

and the general solution of equation (9) assumes the form

$$y = \tilde{y} + Y = A \sin (\omega t + \varphi) - \frac{at}{2\omega} \cos \omega t. \tag{11}$$

It follows from formula (11) that in this case, as in the preceding one, we have a composite oscillatory motion which results from the addition of two oscillations with equal frequency.

The presence of the factor $t$ in the second term suggests that the amplitude of the oscillations increases infinitely with an infinite increase in time $t$, i.e. some time later the load will perform oscillations with a very large amplitude even if the amplitude $a$ of the external force is small. This phenomenon is called *resonance*. It sometimes leads to a destruction of oscillatory systems.

Discussing a load suspended by a spring, we have considered a case of mechanical oscillations of elastic systems (here belong oscillations on the springs of cars, carriages etc). A similar investigation is performed when electric, acoustic and many other oscillations are studied. Linear differential equations play a significant part in these investigations.

It must be pointed out in conclusion that the theory of second-order linear differential equations we have presented can be extended in full to linear differential equations of any order.

# INDEX


Abel, Niels Henrik 438
Abscissa 46, 251
Absolute value of a number 30
Absolutely convergent series 437, 454
Algebraic, adjunct 298
  form of a complex number 451
  polynomial 86, 104
Alternating series 434
Amplitude of oscillations 503
Analytic function 456
Analytical method of representation of a function 84
Angle, between planes 276
    between a straight line and a plane 284
    between straight lines 58, 282
  of inclination of a straight line to the $x$-axis 55
Angular coefficient 56
Antiderivative 181
Applicate 251
Area, of curvilinear sector 227
  of a curvilinear trapezoid 224-226
  of a plane figure 224, 225, 355, 385
  of a stepped figure 224
  of a surface 356-357
    of revolution 232-235
  of a triangle 47
Argument 83, 309
  of a complex number 451
Asymptote 168
  horizontal 168
  of a hyperbola 70
  oblique 168
  vertical 168
Average velocity 123
Axes, of an ellipse 67
  of a hyperbola 71
  of coordinates 46
Axioms of real numbers 23
Axis 25
  of abscissas 46, 251
  of applicates 251
  of ordinates 46, 251
  of a parabola 77

Basic, elementary functions 86
  rectangle of a hyperbola 71
Basis 261
Binomial formula 159
Bolzano, Bernhard 108
Boundary of a domain 316
Boundary point of a domain 316
Bounded, function 83
  sequence 33
    from above (from below) 33
  set 28, 317
    from above (from below) 28

Cantor, Georg Ferdinand Ludwig Philipp 114
Canonical equations, of an ellipse 66
  of an ellipsoid 284
  of an elliptic paraboloid 287
  of a hyperbola 69
  of a hyperbolic paraboloid 288
  of a one-sheet hyperboloid 285
  of a parabola 75
  of a second-order cone 290
  of a two-sheet hyperboloid 286
Cauchy, Augustin Louis 90
Cauchy's conditions 468
  theorem 466, 480
Centre, of an ellipse 67
  of a hyperbola 71
  of mass of a plane 360
    of a body 394
  of a second-order curve 79
Characteristic equation 494
Circle of convergence of a power series 455
Circular cylinder 274
Circulation of a vector field 419
Closed, domain 317
  interval 28
Coefficient, Fourier 460
  of a power series 437
  of a trigonometric series 457
Cofactor 298
Collinear vectors 253
Column 292
  vector 293
Comparison, of infinitely large numbers 100-107
  of infinitesimals 100
  of real numbers 23
  test 429
Composite function 116
Condition for parallelism, of planes 277
    of a straight line and a plane 283
    of straight lines 58, 282
  for perpendicularity, of planes 277
    of a straight line and a plane 283
    of straight lines 58, 282
Conditionally, complex plane 449
  convergent series 437
Conjugate, axis of hyperbola 71
  complex number 451
  hyperbola 71
Connected set 316
Constant 129
  function 83
  number 129
  of a trigonometric series 457
Continuity, of a function, in an interval 106
    at a point 102
    of two variables 314-315
  of real numbers 23
Contour line 311
Convergent, sequence 36, 312, 453
  series 423, 454
Convexity, downwards 164
  upwards 164
Coordinate of a point 27
Coordinate, angle 461
  axis 26
  plane 46, 252
Coordinates of a vector 254
Coplanar vectors 265
Cramer, Gabriel 302
Cramer's rule 302
Criterion of monotonicity of a function 161
Critical point 166
Curve, of the first order 60
  of the second order 64
Curvilinear, cylinder 345
  sector 227
Cycloid 230
Cylindrical, coordinates 392
  surface 273

D'Alembert, Jean le Rond 430
D'Alembert's test 430
Darboux, Jean Gaston 205
Darboux, sum 205
  limit of integration 178
  side of a surface 358
Decreasing, function 117
  sequence 42

Definite integral 202
applications in physics and geometry 224
conditions of existence 204
evaluation 214-216
geometric meaning of 225
properties 212-214
De Moivre, Abraham 452
Density of sources 418
Dependent variable 83, 309
Derivative 121
of a composite function 323
directional 329
finite 121
geometric meaning of 122
of higher order 139
infinite 121
physical meaning of 123
Descartes, René 45
Determinant 295
of second order 269
of a system of equations 301
of third order 296-300
elements of 296
Wronski's 487
Diameter of a domain 344
Direction cosine of a vector 256
Direction of traverse of a contour 371, 400
negative 371, 400
positive 371, 400
Direction field 420
Directional derivative 329
Difference, of complex numbers 450
of real numbers 24
of sequences 33
of vectors 257
Differences 176
Differentiable function 124-125, 319, 366
Differential 126
application in approximations 127-128
of a function of two variables 326-327
geometric meaning of 127, 327
of higher order 333
Differential equation 465
of the first order 465
of forced oscillations 502
of free oscillations 502
of the nth order 484
of the second order 480, 482
with variables separable 469
Differentiation 125
of the basic elementary functions 128, 135
of a composite function 135, 323
of a function represented parametrically 145
principal rules 128, 135
table of derivatives of the basic elementary functions 138
Direct integration 185
Directed, segment 25, 26, 252
tangent 373
Direction, cosines of a vector 255
field 469
Directrices, of an ellipse 72
of a hyperbola 72
Directrix, of a cylindric surface 273
of a parabola 75
Dirichlet, Peter Gustav Lejeune 85
Distance, from a point to a plane 278
from a point to a straight line 283
between two points 47
Divergence of a vector field 416
Divergent, sequence, 37
series 423
Division of a segment in a given ratio 48
Domain 316
of convergence of a power series 437
of definition of a function 83, 309
of integration 344, 366, 396
Dot product 264
Double integral 343-344
calculation by a change of variable 352-354
calculation by reduction to a repeated integral 347-351
geometric applications 355-356
physical applications 359-362
principal properties 340
Eccentricity, of an ellipse 67
of a hyperbola 72
Element, of an area 344
of integration 183
of a matrix 292
of a sequence 32
of a set 21
of volume 366
Elementary, factor 191
function 86
Ellipse 64
Ellipsoid 284
Elliptic, cylinder 274
paraboloid 287
type of a second-order curve 80
Empty set 21
Endpoint 26
Equal, complex numbers 450
vectors 253
Equation, of a curve 53, 273
of an ellipse 66
of an ellipsoid, canonical 284
of an elliptic paraboloid, canonical 287
of forced oscillation of a load 502
of free oscillation of a load 502
of a graph 83
of a hyperbolic paraboloid, normal 288
of an imaginary ellipse 80
of Laplace 422
linear 471
of a one-sheet hyperboloid, canonical 285
of a pair, of imaginary parallel straight lines 62
of coincident straight lines 82
of intersecting straight lines 81
of parallel straight lines 82
of a parabola 75
of a paraboloid cone, normal 290
of a plane, general 275
normal 277
of a straight line, in normal form 62
in slope-intercept form 56
of a surface 272
of a two-sheet hyperboloid, normal 286
with variables separable 469
Equations, of the centre of second-order curve 79
of a curve 273
of a straight line, canonical 280
general 280
Equiangular hyperbola 71
Equilateral hyperbola 71
Equivalent infinitesimals 100
Error 341
Exact differential equation 472
Expansion, of elementary functions into Maclaurin's series 443
of a function into a power series 443
Exponential function 85, 133, 456
of a complex variable 456
Euler, Leonhard 198
Euler's substitution 199

Field 412
Finite, derivative 121
interval 28
limit 26
First remarkable limit 93
First-order curve 60
Flux of a vector across a surface 416
Focal radii of a point 65
Foci, of an ellipse 64
of a hyperbola 68
Focus of a parabola 75
Formula, binomial 159
of Cauchy 151
for a change of variable in a definite integral 221
in an indefinite integral 186
of finite increments 149
of Green 374
for integration by parts in a definite integral 165, 223
in an indefinite integral 196
of Lagrange 149
of Leibniz 140
of Maclaurin 158
of mean value 216
of Moivre 452
of Newton-Leibniz 219
of Ostrogradsky 405
for parabolas 247
of Simpson 247
of Stokes 410
for a substitution in a definite integral 221
of Taylor 155-157, 335
for trapezoids 244
Formulas, of Cramer 302
of Euler 457
for inverse transformation 352
for transformation of coordinates 352
Fourier, Jean Baptiste Joseph 459
Fractional rational function 86, 104
Free vector 253
Frequency of oscillations 503
Function 83
analytic 456
basic elementary 86
table of derivatives of 138
bounded 83
composite 116
constant 104
continuity in an interval 106
decreasing 117
differentiable 124-125, 319, 366
of Dirichlet 85
domain of definition of 83, 309
elementary 86
exponential 86, 133, 456
fractional rational 86, 104
graphical method of representation 86
greatest lower bound of 109
increasing 117
infinitely, large 98
small 96
integrable 203, 344, 364, 388, 398
integral rational 86
inverse 118
trigonometric 86, 134
irrational 86, 196
least upper bound of 109
limit of 87
linear 86
linearly independent 486
logarithmic 86, 132
monotone 117
nondecreasing 117
nonincreasing 117
oscillation of 113
parametric representation 144
piecewise continuous 107
power 86, 104, 130, 138
range of 83, 309
rational 86, 197
scalar 368
of several variables 308
strictly monotone 117
superposition of 116
tabular method of representation 85
transcendental 86, 196
trigonometric 86, 104, 105, 130, 456
uniformly continuous 114
Functional determinant 352
Fundamental identity 26

General, equation of a second-order curve 80
of a straight line 59
solution of a differential equation 467, 481, 484
element of a sequence 32
term of a sequence 32
of a series 423
Generalized mean value theorem 151
Generatrix of a cylindrical surface 273
Geometric meaning, of a definite integral 225
of a derivative 122
of a differential 127, 327
of a differential equation of the form $y' = f(x, y)$ 469
of a double integral 345
of a line integral of the first kind 365
of a triple product 270
Geometric representation of a complex number 449
of a function of two variables 310
o real numbers 25-26
Gradient 330
Graph 84
Graphical method of representation of a function 86
Greatest lower bound of a function 109
Green, George 374

Half-life of a substance 479
Hamilton, Sir William Rowan 420
Harmonic field 422
series 427
Heine, Heinrich Eduard 90
Hermite's interpolation 176-177
Homogeneous linear differential equation 422, 435-439, 471, 485
Hooke, Robert 502
Horizontal asymptote 168
Hyperbola 68
Hyperbolic, paraboloid 288
type of a secondorder curve 80
Hyperboloid, of one sheet 285
of two sheets 286

Identical transformation 295
Identity 53
matrix 294
Imaginary, axis 450
complex number 450
part of a complex number 449
unit 450
Improper integral, of the first kind 237
of the second kind 239
Increasing, function 117
sequence 42
Indefinite integral 183
principal methods of integration 185-190
properties 183-184

Independent variable 63, 309
Indeterminate forms 151
Inequality 23
Infinite, derivative 121
interval 28
Infinitely, large function 98
large sequence 34
small function 96
small sequence 35
Infinitesimal function 96, 314
Initial, conditions 467, 481, 484
phase of oscillations 503
Instantaneous velocity 123
Integrable function 203, 344, 364, 388, 396
Integral, curve 466, 480
part of a number 35
sum 202, 344, 363, 388, 400
test 387, 432
Integrand 183, 203
Integration 183
direct 185
of irrational functions 196-202
by parts 188
of rational functions 191-196
by substitution 186
of total differentials 381-382
of transcendental functions 172
variable 183, 203, 344, 388
Intercept equation of a straight line 60, 61
incomplete 60
Interior point of a set 316
Intermediate variable 116
Interpolation 173
polynomial 173
Interval 28
of conversion of a power series 439
Invariant of the general equation of a second-order curve 80
Inverse, function 118
matrix 305
trigonometric functions 86, 134
Irrational, function 86, 196
number 22

Jacobi, Karl Gustav Jacob 352
Jacobian 352

Lagrange, Joseph Louis 149
Lagrange's interpolation formula 175
Laplace, Pierre Simon 422
Laplace's field 422
Law of the mean 346
Leading diagonal of a determinant 296
Least upper bound of a function 109
Left-hand, derivative 124
limit 90-91
trihedral of vectors 265
Leibniz, Gottfried Wilhelm von 141
Leibniz's test 434
Length, of the arc of a curve 226
of a directed segment 26
of a vector 252
Level line 311
L'Hôpital, Guillaume François Antoine de 151
L'Hôpital's rule 151
Limit, of a function 87
of an integral sum 344
of length of a polygonal line 228
of a sequence 36, 37, 312
of complex numbers 452
Limiting position of a secant 122
Line integral 363
of the first kind 363-365
of the second kind 367, 374
Linear, differential equation 435, 485
equation 471
function 86
Linearly, dependent function 496
independent function 486
Local, extremum 161, 337
maximum 161, 337
minimum 161, 337
Logarithmic, derivative 137
function 66, 132
Logical symbols 22
Lower, bound of a set 28
Darboux sum 205
limit of integration 203
side of a surface 400

Maclaurin, Colin 158
Maclaurin's series 443
Major axis of an ellipse 67
Mapping 63
Mass, of a body 394
of a plate 359
Matrix 292
rectangular 292
square 292
Maximum value of a function 113
Mean value 216
Member of a series 423
Method, of a change of variable 186
of Euler 475
of a fork 176
of integration by parts 188
of least squares 341-343
of Newton 179
of substitution 186
of tangents 179
of undetermined coefficients 192
of variation of a constant 471
Minimum value of a function 113
Minor axis of an ellipse 67
Mixed partial derivatives 331
Möbius, August Ferdinand 399
Möbius strip 399
Modulus of a complex number 451
Moment of inertia, of a body 394
of a plate 361-362
Monotone, function 117
sequence 42
Multiplicity of a root 191
Mutually inverse, functions 120
matrices 306

Negative direction of traverse of a contour 371, 400
Neighbourhood of a point 28, 312, 318
Newton's interpolation formula 176
Nodes of interpolation 173
Nondecreasing, function 117
sequence 42
Nonhomogeneous linear differential equation 471, 485, 495
Nonincreasing, function 117
sequence 42
Nonorientable surface 400
Nonstationary field 412
Nonsteady-state field 412
Normal 61, 277
to a surface 328
Normal vector of the plane 276
Normalizing factor 64
Null vector 252
Number e 43, 160
Number, line 26
sequence 32
series 423
with complex terms 453

o small 100, 101
Oblique asymptote 168

Octant 252
One-sheet hyperboloid 285
One-sided surface 399-400
Open, domain 316
interval 28
set 316
Operator, Hamiltonian 420
Laplacian 422
"nabla" 420
Order of smallness 100
Ordered, pair of numbers 83
triad of vectors 265
triple of numbers 251, 309
Ordinate 46, 251
Orientable surface 400
Orientation of a surface 400
Origin of coordinates 26, 46, 251
Oscillation of a function 113
Ostrogradsky, M. V. 405
Ostrogradsky's theorem 405

Parabola 75
Parabolic type of a second-order curve 80-81
Paraboloid cone 290
Paraboloid of revolution 288
Parallel displacement of coordinate axes 51
Parallelepipedal product of vectors 270
Parameter 144
of a parabola 75
Parameters, of an elliptic paraboloid 288
of a hyperbolic paraboloid 290
Parametric representation of a function 144
Partial, derivative 318
higher-order 331
increment 318
sum of a series 423
Particular solution of a differential equation 467, 481-482
Path of integration 364
Peano, Giuseppe 158
Periodic extension of a function 460
Piecewise continuous function 107
Piecewise smooth, curve 363
surface 395
Point, critical 166
of discontinuity 106, 315
of the first kind 106
of the second kind 106
of division 177
of inflection 165
of local extremum 162, 337
maximum 161, 337
minimum 161, 337
of the number line 28
of a possible extremum 162, 338
of a set 21
singular 166, 239
Point-slope form of the equation of a straight line 57
Polar, angle 49
axis 49
coordinates of a point 49
radius 49
system of coordinates 49
Pole 49
Positive direction of traverse of a contour 371, 400
Potential 413
Potential, field 413
function 414
vector 413
Power, function 86, 104, 130, 138
series 437
with complex terms 454
Principal diagonal, of a determinant 296
of a matrix 294

Product, of complex numbers 450
of matrices 293
o a matrix by a number 293
of real numbers 23
of a sequence by a number 33
of sequences 33
of a vector by a number 257
Projection of a vector on an axis 253
Pure imaginary number 450

Quadrant 46
Quotient, of complex numbers 450
of real numbers 24
of sequences 33

Radius of convergence of a power series 455
Radius vector 291
Range of a function 83, 309
Ratio 48
Rational, function 86, 197
of two variables 197
number 22
Real, axis 450
number 22
part of a complex number 449
Rectangular, coordinates of a point 46, 251
matrix 292
system of coordinates 46, 251
on the plane 46
in space 251
Recursion formula 190
Remainder, of interpolation 177
in Lagrange's form 157
in Peano's form 158
Representation of a vector in terms of base vectors 261
Resonance 506
Right-hand, derivative 124
limit of a function 90
trihedral of vectors 265
Rolle, Michel 148
Rotation, of coordinate axes 51
of a vector field 419
Row 292
vector 239
Rule of triangles 296
Rules of differentiation 128, 135

Scalar, field 412
function 412
product of vectors 262-264
quantity 252
triple product 270
Scheme of investigating the graph of a function 171
Second remarkable limit 95
Second-order, cone 290
curve 64
determinant 269
Secondary diagonal of a determinant 296
Segment 16
Semi-axes of an ellipsoid 285
Semimajor axis of an ellipse 67
Semiminor axis of an ellipse 67
Semiconjugate axis of a hyperbola 71
Semitransverse axis of a hyperbola 71
Sequence 32
of nested intervals 44
Series 423
of complex numbers 449
of Fourier 460
for even and odd functions 461
for functions with period $2l$ 463
of Maclaurin 443
with terms of arbitrary signs 435
Set 21
of values of a function 83, 309

Side of a surface 400
Simple domain 374, 405
Simply connected domain 377
Simpson, Thomas 247
Simpson's formula 247
Singular point 166, 239, 467
Slope of a straight line 56
Smooth, curve 363
surface 395
Solenoidal field 418
Solution of a differential equation 466, 480, 484
Sphere 285
Spherical coordinates 392
Spiral of Archimedes 55
Square matrix 292
Static moment 361
Stationary field 412
Steady-state field 412
Stokes, Sir George Gabriel 408
Strict inequality 23
Strictly monotone, function 117
sequence 42
Subset 21
Sum, of complex numbers 450
of matrices 293
of real numbers 23
of sequences 33
of series 423
of vectors 256
Superposition of functions 116
Surface 399
nonorientable 400
orientable 400
Surface integral 395
of the first kind 395-398
of the second kind 399-405
System of three first-degree equations with three unknowns 300

Table of derivatives, of the basic elementary functions 138
of fundamental integrals 185
Tangent 122
plane 327
Taylor, Brook 155
Term, of a sequence 32
of a series 423
Test, of D'Alembert 430
of comparison 240, 429
for convergence 240
of Leibniz 434
Theorem, of Abel 436, 455
of Bolzano-Cauchy, first 108
second 109
of Cantor 114
of Cauchy 150, 466, 480
on a change of variable in a double integral 352
on comparison of improper integrals 240
of constancy of sign of a continuous function 107
on continuity of a composite function 117
of an inverse function 119
on convergence of improper integrals 240
of monotone bounded sequence 42
on expansion of a rational function into partial fractions 192
on equality of mixed derivatives 332-333
on the general equation of a straight line 56
on independence of a line integral of the integration path 377-380
of L'Hôpital 151
on mean value 216, 346
on monotonicity of a function 161
on nested intervals 44
of Rolle 146
of Taylor 155-157, 335
of Weierstrass, first 110
second 112
Theorems on absolute values 31
on antiderivatives 182, 186, 188
on continuous functions 103, 108-109, 110-118
on convergence, of number series 425-435
of power series 439-444
on convergent sequences 38-41
on differentiable functions 124, 128, 320-323
on the direction of convexity at the inflection points of the graph of a function 165-166
on expansion of functions into Fourier series 459-461
on extrema of functions 162, 337-338
on infinitely small sequences 35
on infinitesimal functions 97
on integrability of functions 204, 208-211, 345
on limits of functions 89-93, 313
on methods of calculation of definite integrals 221, 222
on properties of the ellipse and the hyperbola 73-74
on properties of solutions of linear differential equations 485, 490, 493
on reduction of a double integral to an iterated integral 347-349
on vectors 254-255, 259-262, 264, 269-272
Torus 232
Total differential 377
increment of 315
integration of 361-362
Transcendental function 86, 196
Trapezoid formula 244
Trigonometric form, of a complex number 451
of a function 86, 104, 105, 130, 456
Triple integral 388
applications of 394-395
calculation of 388
by a change of variable 391-392
Two-point form of the equation of a straight line 57
Two-sheet hyperboloid 288
Two-sided surface 399-400

Unbounded sequence 33
Uniformly continuous function 114

Variable of integration 183, 203, 344
Vector 252
field 412
flux 416
function 412
product 265-269
quantity 252
Vertex, of an elliptic paraboloid 288
of a hyperbolic paraboloid 290
of a parabola 77

Weierstrass, Karl Theodor Wilhelm 110
Wronski, Josef Maria Hoené 487
Wronskian 487

Zero 450
Zero vector 252
Zero-divergence field 418